Causal modelling in epidemiology and health services research

Zoe Fewell

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Medicine and Dentistry

Social Medicine

January 2007

Word count: 70,576

Abstract

Measurement error in explanatory variables and unmeasured confounding can lead to biased estimates of exposure-outcome associations in observational epidemiological studies. In this thesis, simulation studies show that the effects of residual and unmeasured confounding on the estimated exposure-outcome association can be complex, especially when the confounders are correlated. Unmeasured confounding is a greater problem when the confounders are uncorrelated. In addition, classical measurement error in the exposure variable attenuates the estimated association towards the null.

Two methods for correcting for measurement error in explanatory variables, regression calibration and simulation-extrapolation, are investigated via simulation study. The regression calibration method performs well when there are no unmeasured confounders, removing all of the bias due to measurement error in the explanatory variables. The simulation-extrapolation method reduces bias but does not remove it completely.

A sensitivity analysis is used to investigate residual confounding by triglycerides and forced expiratory volume in one second (FEV_1), and exposure measurement error in C-reactive protein (CRP) on the association between CRP and coronary heart disease in the British Women's Heart and Health Study. Allowing for measurement error in FEV_1 has little impact on the estimated exposure-outcome association. Allowing for measurement error in triglycerides slightly attenuates the estimated association, while allowing for measurement error in CRP moves the estimated association away from the null value.

Non-compliance in randomised controlled trials can result in an intention-to-treat analysis that estimates the effectiveness of treatment, but not its efficacy. Data from the Nambour Trial are used to investigate the effect of sunscreen use on time to basal cell carcinoma accounting for compliance. A rank preserving structural failure time model is used for the analysis, as it allows failure time outcomes and complex compliance data. The compliance analysis moves the point estimate away from the null value when compared with the intention-to-treat estimate.

Acknowledgements

I would like to acknowledge the help and advice of my supervisors, Prof. Jonathan Sterne and Prof. George Davey Smith. Dr. Kate Tilling also provided temporary supervision in the early stages of my PhD, for which I am extremely grateful.

Two of the chapters in this thesis are analyses of real life data. Thanks therefore also go to Prof. Debbie Lawlor and the British Women's Heart and Health Study team for allowing me to use their data in my investigation of the impact of residual confounding and exposure measurement error in observational studies, presented in Chapter 7. Many thanks also go to Prof. Adèle Green, Prof. Gail Williams and the Nambour Skin Cancer and Actinic Eye Disease Trial team for allowing me to use data from the Nambour Trial in my investigation of the effect of compliance in a randomised controlled trial, presented in Chapter 10. Chapter 10 was also much improved by the helpful comments provided by Dr. Ian White.

Mum, Dad, Chris and Jon have all been supportive throughout the research and writing process of this Ph.D. I would like to thank all of them for listening when things were difficult, and for encouraging me to finish it.

This work was funded by a research studentship, provided by the Medical Research Council. I am extremely grateful for this financial support.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of the dissertation has been submitted for any other academic award. Any views expressed in the dissertation are those of the author.

SIGNED: Zoe Fewell DATE: 09/02/07

Table of contents

List of tables

List of figures

List of abbreviations

ARIC	Atherosclerosis Risk In Communities
BC_a	Bias-corrected and accelerated
BCC	Basal cell carcinoma
BMI	Body mass index
BWHHS	British Women's Heart and Health Study
CACE	Complier average causal effect
CHD	Coronary heart disease
CHS	Cardiovascular Health Study
CI	Confidence interval
C-PROPHET	Complier's proportional hazards effect of treatment
CRP	C-reactive protein
CV	Coefficient of variation
CVD	Cardiovascular disease
EEE	Expected estimation equations
EM	Expectation-maximisation
FEV_1	Forced expiratory volume in one second
HDL	High density lipoprotein
Health ABC Study	Health, Aging and Body Composition Study
HPFS	Health Professionals Follow-up Study
HRT	Hormone replacement therapy
ICC	Intra-class correlation coefficient
IHD	Ischaemic heart disease
ITT	Intention-to-treat
LATE	Local average treatment effect
LDL	Low density lipoprotein
LRC-CPPT	Lipid Research Clinics Coronary Primary Prevention Trial
MI	Myocardial infarction
MLE	Maximum likelihood estimator
MONICA	Monitoring Trends and Determinants in Cardiovascular Disease
MRFIT	Multiple Risk Factor Intervention Trial
mWh cm^{-2}	Milliwatt hours per centimetre squared
Nambour Trial	Nambour Skin Cancer and Actinic Eye Disease Trial
NHS	Nurses' Health Study
NMSC	Non-melanoma skin cancer
OR	Odds ratio
PHS	Physicians' Health Study

PRIME	Prospective Epidemiological Study of Myocardial Infarction
RCT	Randomised controlled trial
RPSFTM	Rank preserving structural failure time model
RR	Relative risk
SCC	Squamous cell carcinoma
SimEx	Simulation-extrapolation
SNMM	Structural nested mean model
SUTVA	Stable unit-treatment value assumption
UV	Ultraviolet
UVB	Ultraviolet-B
WHS	Women's Health Study

Chapter 1.

Introduction

1.1. Causal effects in epidemiology

The aim of epidemiologic analysis is often to estimate the causal effect of an exposure on an outcome of interest. The problems with using observational studies for causal inference are well known.[1, 2] Selection bias, recall bias, loss to follow-up and reverse causation are some of the problems that can lead to biased estimates of associations between exposure and outcome in observational studies. Confounding is caused by variables associated with both outcome and exposure, and not on the causal pathway between exposure and outcome. Controlling for variables with these properties may remove bias, but they must all be measured perfectly, and their association with the exposure of interest perfectly characterised. There are also variables for which control does not remove bias even though they have all three properties of confounders.[3] While it is recognized that, under certain conditions, non-differential measurement error in the exposure leads to a bias towards the null,[4] the effects of measurement error in confounders, which leads to residual confounding, are not well understood.

Consider, for example, the effect of antioxidant vitamin intake on cancer, cardiovascular disease, and mortality. Observational studies have shown protective effects,[5-7] while in contrast randomised trials have shown no effect on outcomes.[8, 9] It has been suggested that the disparity in results is likely to be due to confounding by behavioural and social factors acting across the life course.[10, 11] For example, factors related to childhood social class may be important confounders for the association between antioxidant vitamin intake and disease outcome. To fully capture the life course effect of such confounders, all related factors must be measured perfectly. Failure to do so, either due to unmeasured confounders or residual confounding, will result in biased estimates of the causal exposure effect. There are, of course, other possible explanations for this disparity in results. Observational studies estimate the effect of dietary exposure on disease outcomes, whereas randomised trials change exposure in the intervention group. This change is often by supplementation, and can result in exposure to the nutrient of interest much higher than would ever be seen in an observational study. In some situations, however, the exposure in the observational studies is precisely the same as the treatment in the randomised controlled trials (RCTs). For example, two widely cited papers in 1993 demonstrated substantially lower coronary heart disease (CHD) risk amongst people using vitamin E supplements, that were apparently robust to confounding.[6, 7] Furthermore, there was no trend towards increasing protection by vitamin E supplements when taken for more than two years. The exposure in these observational studies is precisely the same exposure as that tested in RCTs of vitamin E supplementation that have run for up to six years. In these RCTs there is robust evidence of no material effect on CHD risk.[12]

Some authors debate whether residual and unmeasured confounding can cause large exposure-outcome effect estimates. Morabia[13] asserted that strong associations are unlikely to be

completely attributable to confounding because strong confounders are likely to be detected in the study population, or recognized in the literature as strong confounders and therefore measured and controlled for in the analysis. More recently, Khaw, Day, Bingham *et al.*[14] advanced similar arguments with respect to whether residual confounding could explain observed associations between plasma ascorbic acid and mortality. Much of the literature describing the effects of residual confounding on exposure effect estimates assumes that there is a single confounder of the exposure-outcome association. In these cases, the assertions made by Morabia[13] may generally (but not always) apply.

When measurement error in explanatory variables exists, statistical methods have been proposed to correct exposure effect estimates for the bias caused by measurement error. These methods involve using external or internal validation studies, replicate measures or transportation to estimate the measurement error parameters, such as the error variance for continuous variables or the sensitivity and specificity for binary variables. Sensitivity analysis methods can be used in situations in which the measurement error or misclassification parameters cannot be estimated from available data, by using a range of plausible values for the measurement error parameters to show the possible impact of measurement error on estimated exposure effects.

It may also be difficult to estimate causal treatment effects from RCTs. The primary method of analysis in RCTs is by intention-to-treat (ITT). In an ITT analysis, participants are analysed based on their randomly allocated treatment. An ITT analysis estimates the effectiveness of treatment, i.e. the effect of a treatment policy. Another estimate of interest is the efficacy of treatment, i.e. the effect of receiving treatment. If participants do not comply with their allocated treatment, the ITT analysis does not estimate of the efficacy of treatment. The ITT estimate is instead a combination of the effect of treatment among compliers, and the lack of effect among non-compliers. As-treated and per-protocol analyses generally produce biased efficacy estimates. As-treated analyses, in which the randomisation is ignored and participants analysed based on treatment actually received, are subject to the same problems as observational analyses. Per-protocol analyses, in which subjects are censored at the first time they depart from their randomised treatment, may be subject to selection bias. Statistical methods have been developed to analyse data from RCTs with non-compliance that respect the randomisation and provide an estimate of treatment efficacy. These analysis methods are preferable to as-treated or per-protocol analyses when an estimate of treatment efficacy is required.

1.2. Aims

This thesis is divided into three distinct parts. In Part A, the effects of residual and unmeasured confounding and exposure measurement error on exposure effect estimates are considered.

Simulation studies demonstrate the effects of such errors on exposure-outcome ORs estimated from logistic regression models. In Part B, some of the results obtained from the simulated datasets of Part A are corrected for measurement error using two methods; regression calibration and simulation-extrapolation. A sensitivity analysis is performed using data from the British Women's Heart and Health Study, and the effect of residual confounding and exposure measurement error on the estimated association between C-reactive protein (CRP) and CHD is investigated. Part C considers methods for analysis of data from RCTs when participants depart from their randomly allocated treatment and demonstrates the use of one particular model, a rank preserving structural failure time model, in an RCT investigating the effect of sunscreen use on basal cell carcinoma (BCC), where data on sunscreen used are available.

The aims of this thesis are as follows:

1. To review the literature on the bias in the causal exposure effect estimate caused by exposure measurement error and misclassification in observational epidemiological studies (Chapter 2).
2. To review the literature on the bias in the causal exposure effect estimate caused by residual confounding in observational epidemiological studies (Chapter 2).
3. To review the literature on the bias in the causal exposure effect estimate caused by unmeasured confounding in observational epidemiological studies (Chapter 2).
4. To investigate the impact of residual and unmeasured confounding in logistic regression analyses with either two or four confounders using a simulation study (Chapter 3).
5. To investigate the impact of exposure measurement error, in addition to residual and unmeasured confounding, in logistic regression analyses with either two or four confounders using a simulation study (Chapter 3).
6. To review the literature on methods available to correct for measurement error in observational studies in cases where the amount of measurement error is known or can be estimated (Chapter 4).
7. To review the literature on sensitivity analysis methods in epidemiology (Chapter 4).
8. To use simulation studies to investigate two methods for correcting for measurement error, regression calibration and simulation-extrapolation, using datasets described in Chapter 3 (Chapter 5).
9. To review the literature on the association between CRP and CHD (Chapter 6).
10. To investigate the impact of residual confounding and exposure measurement error on the association between CRP and CHD using data from the British Women's Heart and Health Study (Chapter 7).
11. To review the literature on methods available to estimate treatment efficacy in RCTs with departures from randomly allocated treatment (Chapter 8).
12. To review the literature on the association between sunlight and BCC (Chapter 9).

13. To use rank preserving structural failure time models to estimate treatment efficacy, allowing for departures from randomly allocated treatment, in an RCT of the effect of sunscreen use on time to first BCC using data from the Nambour Skin Cancer and Actinic Eye Disease Trial (Chapter 10).

BLANK IN ORIGINAL

Part A.

The effects of measurement errors in explanatory variables and unmeasured confounding on estimated exposure-outcome associations

BLANK IN ORIGINAL

Chapter 2.

Background literature: The effects of measurement error and misclassification in explanatory variables and unmeasured confounding in epidemiological studies

2.1. Introduction

In this chapter, literature describing the effects of measurement errors in exposure or confounding variables, or omission of confounders on estimated exposure effects is described. This is motivated by an investigation of the impact of unmeasured confounding, caused by omitting confounders from a regression analysis, residual confounding, caused by measurement error in confounders, and exposure measurement error on estimated exposure effects from logistic regression models, which is presented in Chapter 3. First, types of measurement error and measurement error models are described. This is followed by a review of the literature describing the effects of measurement errors in exposure variables on exposure effect estimates. The next section describes the effect of measurement errors in confounding variables on exposure effect estimates. The final section of this chapter describes the effect of unmeasured confounding on estimated exposure-outcome effects.

2.2. Types of measurement error and measurement error models

Throughout this chapter, X will denote a variable measured without error. This will be referred to as the *true value* of the variable. In the context of epidemiological studies, and measurement error in an exposure or confounder, the true value is defined to be the value that is causally related to the outcome of interest. For example, the long-term average of blood pressure may be causally related to risk of coronary heart disease, but it is difficult to measure. A single blood pressure measurement is unlikely to capture the long-term average accurately, due to natural fluctuations in blood pressure, calibration error in the measuring instrument, or a person's anxiety level. This single measurement can be regarded as a measurement of the long-term average blood pressure which is subject to measurement error. Throughout this chapter, measurements that are subject to error will be denoted by Z. The effects of measurement errors depend on the type of measurement error, and how it is related to the true value of interest. Types of measurement errors and measurement error models are described below.

2.2.1. Random and systematic measurement error

Random and systematic measurement errors are affected by the value of the variable of interest, and not by other variables that may be related. For continuous variables, random errors can be thought of as adding noise to the measurement of a variable. They do not influence the data in any particular direction. More specifically, the expected value of the error is zero when the error is random.

Systematic errors, on the other hand, will bias the measurement of a variable in a specific direction (i.e. the expectation will be non-zero). Systematic errors can occur, for example, if

laboratory measurements become more variable the closer to the lower limit of detection the measurement lies.

2.2.2. Differential and non-differential error

Differential measurement error occurs when the measurement error depends on the value of another variable. For example, the measured value of the variable may depend on the outcome. This form of measurement error can occur in, for example, case-control studies in the form of recall bias. A case may be more likely to remember and report past exposures than a control, which leads to error in the exposure that is differential with respect to the outcome.

2.2.3. Additive and multiplicative measurement error

Measurement errors can act either additively or multiplicatively on the true value of the variable being measured. If X denotes the true value of the variable, Z denotes the mismeasured variable, and ε is the error, then additive measurement error takes the following form:

$$Z=X+\varepsilon,$$

while multiplicative measurement error takes the form:

Equation 2.1: $$Z=X\varepsilon$$

Throughout this chapter, and the remainder of the thesis, measurement error will be assumed to be additive. It is useful to note that, by taking logarithms of Equation 2.1, the measurement error structure becomes additive:

$$\ln Z=\ln X+\ln\varepsilon,$$

and therefore all characteristics of additive measurement error apply to multiplicative error on a log scale.

2.2.4. Classical and Berkson measurement error models

In the classical measurement error model, the error is assumed to be independent of the true value of the variable. Therefore, if the measurement error is additive, the model for classical measurement error is:

$$Z=X+\varepsilon,$$

and $Var(Z)=Var(X)+Var(\varepsilon)$. A classical measurement error model is appropriate if there is a true level of exposure but it cannot be precisely measured, for example a person's exposure to air pollution.

In the Berkson measurement error model, measurement error is assumed to be independent of the measured values of the variable, and the true value is centred around a proxy measurement. Assuming measurement error is again additive, the Berkson model is:

$$X=Z+\varepsilon,$$

and $Var(X)=Var(Z)+Var(\varepsilon)$. The Berkson measurement error model is appropriate if the dose of

an exposure can be controlled, but the true level of exposure X (e.g. the amount of drug absorbed into the bloodstream) varies randomly around the fixed dose Z.

2.3. The impact of measurement error in exposure

In this section, the effects of measurement error or misclassification in the exposure variable are discussed. Literature on the effects of measurement error or misclassification of confounders is described in Section 2.4. Methods for correcting for measurement error or misclassification will be considered in Chapter 4.

2.3.1. Binary exposure

Early interest in the effects of exposure measurement error focused on 2x2 tables. Bross[15] showed in this setting that non-differential misclassification results in an underestimation of the difference in proportions of cases between two populations (an important correction to the results in this paper is provided by Newell[16]). It was also shown that misclassification reduces power to detect effects. This could be viewed effectively as a loss of study subjects. A similar idea was advanced by Phillips and Davey Smith,[17] who showed that, in some situations, it may be better to recruit fewer people to a study and use study resources to obtain repeated measurements on those subjects than to recruit more people and measure exposures only once.

Greenland[18] showed that non-differential misclassification of a binary exposure variable in the presence of covariates results in attenuation of the observed exposure effect estimate. Considering odds ratios (ORs) in strata defined by values of a binary confounder, non-differential misclassification of the exposure may introduce heterogeneity into the stratum-specific ORs when in truth they are homogeneous, or hide heterogeneity if it truly exists.

The effects of non-differential and random misclassification of a binary exposure variable on estimates of relative risk in the 2x2 table setting was considered by Flegal, Brownie and Haas.[19] A formula was provided that related the observed relative risk to the sensitivity and specificity of the misclassified exposure, the prevalence of exposure, and the true relative risk. This showed that:

1. If the true relative risk equals one, then the observed relative risk equals one and there is no bias, regardless of the sensitivity, specificity or prevalence of exposure.
2. If the sum of the sensitivity and specificity equals one, the observed relative risk will be one regardless of the true relative risk.
3. If the sum of the sensitivity and specificity is less than one, the observed relative risk will show an effect in the opposite direction to the true relative risk.
4. If the sum of sensitivity and specificity is greater than one, the observed relative risk will underestimate the true relative risk.
5. If the sensitivity is perfect (i.e. equal to one), the observed relative risk increases as the

prevalence increases.

6. If the specificity is equal to one, the observed relative risk decreases as the prevalence increases.
7. When both sensitivity and specificity are imperfect, the relationship of the observed relative risk with prevalence of exposure is not monotonic.
8. When comparisons of observed relative risks are made between groups, such as those defined by levels of a confounder, the differences in the prevalence of exposure in the groups can cause heterogeneity between the observed relative risks when there is none in the true relative risk, or mask true heterogeneity.
9. Spurious trends with other variables can also be caused or masked if the prevalence of exposure increases as the third variable increases.

Marshall and Hastrup[20] considered misclassification in a binary exposure and confounder where the errors were uncorrelated. They demonstrated, through simulation studies, that increasing the misclassification in the exposure biases the estimated OR towards the null if the confounder misclassification is held fixed. Additionally, if there is no misclassification in the confounder, and no true association between exposure and outcome, misclassification of the exposure does not introduce any bias into the estimated exposure-outcome OR.

Davidov, Faraggi and Reiser[21] considered misclassification of a binary exposure in a logistic regression setting. Exact and first order approximations of the effect of misclassification were provided. In the general setting, where misclassification may be differential, the bias was shown to depend on the sensitivity and specificity of the misclassification, and the odds of exposure for the two different levels of outcome.

The effects of non-differential misclassification of binary exposure variables have been well discussed in the literature, but often appear to be misunderstood. It is often stated that the exposure effect estimate is biased towards the null when the misclassification is non-differential. In fact, this is not a sufficient condition to guarantee a bias toward the null. The rule in reality applies to the average effect of non-differential measurement error in repeated studies that vary only randomly from each other. Therefore, in a single study, the bias due to non-differential misclassification of an exposure can be either towards or away from the null. This has been demonstrated in simulation studies by Jurek, Greenland, Maldonado *et al.*[4] and Sorahan and Gilthorpe.[22]

2.3.2. Polytomous exposure

Non-differential misclassification of a polytomous exposure was considered by Fung and Howe.[23] Using simulation studies, they showed that non-differential misclassification of exposure results in a bias towards the null value. In addition, exposure misclassification

influences the p-value of a test and results in a loss of power.

Dosemeci, Wacholder and Lubin[24] investigated the effect of non-differential misclassification of polytomous exposures. In contrast to the results of Fung and Howe,[23] they found that misclassification could bias the exposure effect estimate towards or away from the null value, and that in some situations the bias could act to reverse the direction of the estimated effect. Gilbert[25] suggested that, although the misclassification was non-differential, it was also systematic, and that the reversal of trend was due to the systematic misclassification. This proposition was investigated by Weinberg, Umbach and Greenland.[26] They found that a change in direction of the exposure effect estimate could occur if systematic exposure misclassification was present. Furthermore, even if systematic exposure misclassification occurs, if the error can be assumed to be monotonically increasing with the value of the true exposure, a reversal in trend cannot occur.

Dose-response trends are often considered in epidemiological studies with polytomous exposures. Brenner[27] considered the effect of non-differential misclassification of a three-level polytomous exposure on the estimated dose-response relationship. Misclassification between adjacent groups only was considered. Misclassification of the exposure can suggest a trend where there in reality is none. If all levels of the exposure are misclassified (e.g. from low to medium and vice versa, and from medium to high and vice versa), the effects of each misclassification may cancel each other out partially or completely. In situations with a non-monotonic dose-response relationship, misclassification may result in under- or over-estimation of the relationship. In the situation of a monotonic non-linear dose-response relationship, misclassification leads to an attenuation of the estimated trend. P-values are often provided when the trend is assessed. Brenner[27] additionally showed that non-differential exposure misclassification can either increase or decrease the true p-value from a Mantel extension test for trend. An increased p-value is more likely when the true dose-response relationship is monotonic, and a decreased p-value more likely when the true dose-response trend is non-monotonic.

Wacholder, Dosemeci and Lubin[28] showed that collapsing a categorical variable into fewer categories can result in differential misclassification, even if the original variable is non-differentially misclassified. This could occur if, for example, people are categorised as never smokers, ex-smokers, and current smoker subject to non-differential misclassification. In an analysis comparing ever-smokers with never-smokers, the two categories may be subject to differential misclassification. This can occur if the categories that are collapsed have different risks of disease and different probabilities of misclassification. If differential misclassification does arise from collapsing categories, exposure effect estimates may be larger than the true value, or suggest an effect in the opposite direction.

A common practice in epidemiology is to categorise continuous variables. If the continuous variable is subject to non-differential measurement error, the categorical variable that arises from this process may be differentially misclassified.[29] For example, subjects may be more likely to be misclassified the closer to the cutpoint their true exposure value lies. If a higher level of exposure leads to an increased probability of disease then differential misclassification of the exposure variable can occur. In their simulation studies, Flegal, Keyl and Nieto[29] showed that the bias due to differential misclassification arising from categorising a continuous variable may be smaller than if the error was non-differential. Gustafson and Le[30] also showed that the differential misclassification that may arise by dichotomising a continuous variable can result in a smaller bias in the effect estimate than the bias caused by the non-differentially mismeasured continuous variable. Whether this reduction in bias occurs depends on the underlying relationship between the outcome and continuous exposure. It was also suggested that if dichotomising a continuous variable improves model fit, then the bias due to the resulting misclassification will be worse than if the continuous variable was used in the regression.

2.3.3. Continuous exposure

The idea of considering the effects of measurement errors in exposures was first introduced by Spearman in 1904,[31] who noted that errors in continuous variables would attenuate the correlation between the variables.

Measurement error in a normally distributed exposure was considered by Armstrong, Whittemore and Howe.[32] Assuming a discriminant analysis model and a case-control setting, formulae were provided for the bias in the estimated logistic regression coefficients. The formulae account for both random and differential measurement errors. If there is no confounding and no differential error, the exposure effect estimate is attenuated towards the null by a factor of $\sigma_x^2/(\sigma_e^2+\sigma_x^2)$, where σ_x^2 is the variance of the true exposure, and σ_e^2 is the variance of the measurement error. If a confounder is included in the model and measured without error, the attenuation due to measurement error in the exposure is increased. This increased attenuation also occurs when an additional explanatory variable that is not a confounder is unnecessarily included in the analysis.

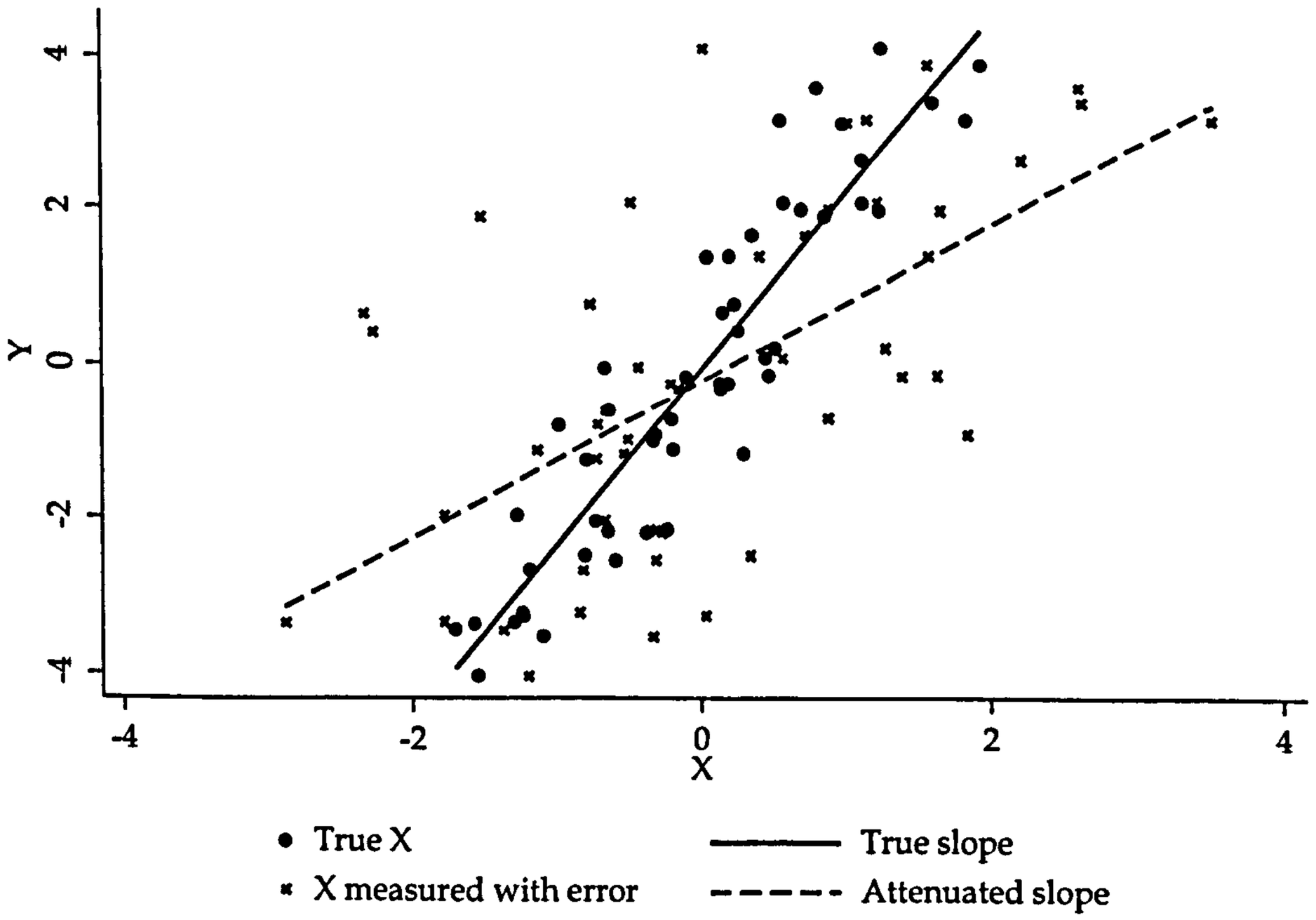
The attenuation factor defined in the previous paragraph is the same as the attenuation in a linear regression when the exposure is measured with error. This effect can be easily demonstrated. The univariable linear regression equation is

$$Y=\beta X+\varepsilon.$$

Suppose that $\beta=2$. A scatter plot of Y against X will have a slope of 2, as shown in Figure 2.1. Now suppose that X is measured with error. Figure 2.1 shows that the scatter plot of Y against the mismeasured X has a greater horizontal spread, which results in a regression slope with smaller gradient.

Armstrong[33] discussed the effect of measurement error in normally distributed exposures, assuming a classical measurement error model with non-differential and random measurement error. The same attenuation factor of $\sigma_x^2/(\sigma_e^2+\sigma_x^2)$ as found by Armstrong, Whittemore and Howe[32] was found for exponential models, or cohorts in which the time to event is observed. The effect of Berkson type measurement error was also considered, and resulted in no attenuation of the relative risk estimate. There was, however, an increase in the standard error of the estimate. These results apply to case-control studies in which cases and controls are matched by age, or cohort studies in which the time to the outcome event is observed. If the time to event is not observed, the results described here do not apply. However, if the risk of disease is small, the effects of measurement error will be close to that described above. Attenuation of the exposure-outcome relationship, and a lack of attenuation in the Berkson measurement error model, have also been demonstrated by Clayton and Gill[34] in the context of linear regression with a normally distributed exposure.

Figure 2.1: Attenuation of linear regression when the independent variable is measured with error.



The effect of measurement error correlated with the true value of an explanatory variable was considered by Wacholder.[35] Assuming a univariate linear regression model, this type of measurement error in the exposure can result in inflation, attenuation or a reversal of direction of the true exposure effect estimate. If the error variance is less than the variance of the true variable, and there is a negative correlation between the error and the true variable, exaggeration of the true effect can occur. Reversal of effects can occur if the error variance is greater than the variance of the true variable, and the correlation between the error and the true

variable is negative. The correlation between exposure and error can lead to a Berkson measurement error model, in which the error is uncorrelated with the exposure as measured. This situation results in no bias in the estimated exposure effect when using the measured exposure as opposed to the true exposure.

In a simulation study with continuous exposure and single confounder, Marshall and Hastrup[20] also showed that, if the errors in the exposure and confounder are uncorrelated, measurement error in the exposure attenuates the estimated exposure effect towards the null for both linear and logistic regressions. This effect occurs regardless of the bias induced by measurement error in the confounder. It was additionally demonstrated that, if the exposure has no true association with the outcome and there is no measurement error in the confounder, exposure measurement error does not induce any bias in the estimated exposure-outcome association. Marshall, Hastrup and Ross[36] showed that bias in the exposure effect estimate can be in either direction if the error in exposure is correlated with the error in confounder in both linear and logistic regressions, depending on the correlation between the errors and the amount of error in the confounder.

Kipnis, Freedman, Brown *et al.*[37] provided a formula for the effects of measurement error in a continuous exposure for linear regression analyses. This formula showed that, under the classical measurement error assumptions of errors uncorrelated with the true variable or each other, measurement error in the exposure leads to an attenuation towards the null of the true exposure-outcome association. The formula also allowed for the classical measurement error assumptions to be relaxed. In these situations, error in the exposure can lead to an observed association in the opposite direction to the true exposure-outcome association, or a larger observed association than the true association.

2.3.4. Summary

The effects of exposure measurement error on estimated exposure effects are summarised below.

- If the measurement error in a continuous exposure is non-differential and random the exposure effect estimate is attenuated towards the null when compared with the true effect.
- If there is no true association between exposure and outcome, non-differential and random exposure measurement error has no effect on the observed exposure effect estimate.
- For non-differential and random misclassification of a binary exposure, the observed relative risk equals one if the sum of the sensitivity and specificity of misclassification equals one.
- For non-differential and random misclassification of a binary exposure, the observed

relative risk is in the opposite direction to the true relative risk if the sum of the sensitivity and specificity of misclassification is less than one.

- For non-differential and random misclassification of a binary exposure, the observed exposure effect is an underestimate of the true exposure effect if the sum of the sensitivity and specificity of misclassification is greater than one.
- Systematic exposure measurement error may bias the observed effect estimate towards or away from the null value, or result in an observed effect in the opposite direction to the true effect.
- Non-differential and random exposure measurement error causes a loss of power to detect effects.
- Non-differential measurement error may introduce or conceal heterogeneity in stratum specific exposure effect estimates.
- When trend is being assessed, exposure measurement error can suggest a trend where there is none.
- Exposure measurement error may result in underestimation or overestimation of a dose-response relationship when the true relationship is non-monotonic.
- Exposure measurement error causes underestimation of dose-response relationships if the true relationship is monotonic.
- The p-value for trend may be overestimated or underestimated when exposure is measured with error.
- The bias induced by collapsing a categorical or continuous variable subject to non-differential measurement error into fewer categories may be smaller than the bias caused by the original variable.
- Including perfectly measured confounders or unnecessary additional variables in a regression model increases the attenuation caused by exposure measurement error.
- Berkson error in the exposure results in no attenuation of the observed exposure effect estimate, but increases the standard error of the effect estimate.
- If the exposure measurement error is correlated with the true exposure variable, the observed exposure effect estimate may be inflated, attenuated or reversed in comparison with the true effect.
- If errors in variables are correlated, exposure measurement error may bias effect estimates in either direction.

2.4. The impact of measurement error in confounders

In this section, the literature describing the impact of measurement error or misclassification in confounders is discussed.

2.4.1. Binary confounders

It is well documented that measurement error in confounders reduces the ability to control for confounding in the analysis. For example, Greenland[18] illustrated the effect of confounder misclassification in the simple setting of dichotomous outcome, exposure and single confounder. Non-differential misclassification of a binary confounder resulted in an OR intermediate between the crude OR and the true OR. The estimated OR could be biased either towards or away from the null value when a binary confounder was misclassified. Considering ORs in strata defined by values of the confounder, it was shown that non-differential misclassification may introduce heterogeneity into the stratum-specific ORs when in truth they are homogeneous, or hide heterogeneity if it truly exists. Walker and Lanes[38] provided a method to assess whether observed heterogeneity in stratum specific ORs is caused by either non-differential or differential misclassification of a confounder.

The effect of dichotomising a continuous confounder was demonstrated by Becher.[39] By simulating a case-control study with a continuous exposure and a binary confounder obtained by dichotomising the true continuous confounder at various cutpoints, it was demonstrated that the bias in the estimated exposure effect increases with increasing correlation between exposure and confounder, and is minimised when the cutpoint is at the overall mean of the continuous confounder when cases and controls are combined. The bias in the estimated exposure effect was also shown to reduce for an increasing number of categories created from the true continuous variable.

An expression for the bias caused by misclassification of a binary confounder on the exposure effect estimate in logistic regression with binary explanatory variables was provided by Davidov, Faraggi and Reiser.[21] An exact formula for the bias in the exposure-outcome log OR was provided, and a Taylor series expansion provided a first order approximation that depended on the sensitivity of the misclassification, but not the specificity:

$$\text{Bias} = (\xi_{01} - \xi_{00})(1-\theta_0) + (\xi_{10} - \xi_{11})(1-\theta_1)$$

In this formula, ξ_{ij} is the retrospective odds of the confounder, when the outcome equals i and the exposure equals j, and θ_i is the sensitivity of the confounder misclassification given the outcome equals i. This suggests that, if θ_0 and θ_1 both equal one, there is no bias in the estimated exposure effect. The direction of the bias in the exposure effect estimate is dependent on $\xi_{01} - \xi_{00}$ and $\xi_{10} - \xi_{11}$. The bias can therefore act to increase or decrease the estimated exposure effect. The results cannot be extended to the case with continuous covariates, as closed form expressions do not exist in this situation for logistic regression models.

2.4.2. Polytomous confounders

Using simulation studies, Fung and Howe[23] demonstrated the effects of misclassification of a

polytomous confounder on exposure-outcome relative risk estimates. An exposure and single confounder, each with four categories, were simulated and subjected to non-differential misclassification. They showed that the bias caused by confounder misclassification may be either towards or away from the null value. The exposure effect estimate was biased away from the null when the association between the confounder and exposure was in the same direction as the association between the confounder and outcome. The exposure effect estimate was biased towards the null when the association between the confounder and exposure was in a different direction to the association between the confounder and outcome. Confounder misclassification also influences the estimated power of the Mantel trend test. When the association between the confounder and exposure is in the same direction as the association between the confounder and outcome, confounder misclassification increases the estimated power, while if the associations are in opposite directions the estimated power decreases.

As stated above, Greenland's[18] results suggested that, for non-differential misclassification of a binary confounder, the adjusted OR is intermediate between the crude OR and the true OR. Brenner[40] showed that this relationship does not necessarily exist for non-differential misclassification of polytomous confounders. Estimated associations can be biased away from the true value in the opposite direction to the crude value, or the estimated association can be further from the true value than the crude value. The misclassification probabilities used to demonstrate these effects resulted in systematic misclassification of the confounder, as the misclassification probabilities were not the same between all categories of the polytomous confounder.

2.4.3. Continuous confounders

The effects of measurement error in continuous confounders were demonstrated by Armstrong[33] in the context of a continuous exposure and a single confounder. A classical measurement error model was assumed, in which the explanatory variables were drawn from a multivariate normal distribution with normally distributed errors that were uncorrelated with either the true covariate values or other error variables. Under these circumstances, measurement error in a continuous confounder results in an exposure effect estimate that is intermediate between the true value and the crude estimate obtained by omitting the confounder from the analysis. This was shown by Greenland[18] in the case of a dichotomous exposure and confounder.

Measurement error in a continuous exposure and confounder was considered by Phillips and Davey Smith.[41] Both the exposure and confounder were subject to random, non-differential measurement errors that were uncorrelated. Bias in the exposure effect estimate, obtained by logistic regression, increased with increasing correlation between the explanatory variables. It was possible for the estimated exposure-outcome association to show an effect in the opposite

direction to the true effect when the exposure and confounder were correlated. These effects increased with increasing correlation between the explanatory variables and with increasing measurement error in the confounder. This suggests that measurement error in confounders is a potentially serious problem, and can lead to effect estimates that erroneously suggest harm or benefit from exposure.

The effect of errors correlated with the true values of continuous confounders was considered by Wacholder[35] in the context of linear regression. It was shown that the exposure effect estimate may be biased away from the null value when the covariance between the error and true confounder is negative and greater in absolute value than the variance of the true confounder. It was additionally required that the relationship between the true confounder and exposure is the same as the relationship between the measured confounder and exposure, and that the error is non-differential.

Uncorrelated measurement errors in the exposure and a confounder were considered by Marshall and Hastrup.[20] Three cases were considered in which the outcome, exposure and confounder were either continuous or dichotomous. In the setting where the exposure had no causal effect on the outcome, error in the confounder biased the estimated exposure-outcome association and introduced an apparent effect. The bias in the estimated exposure-outcome effect was shown to increase with increasing error in the confounder, with the correlation between the exposure and confounder, and with the true association of the confounder with outcome.

An extension of the expression for the effect of measurement error in a confounder on the exposure effect estimate given by Armstrong[33] was provided by Kipnis, Freedman, Brown *et al.*[37] for a linear regression model. It was again assumed that there was a single confounder of the exposure-outcome association. The formula was not restricted by the assumptions of a classical measurement error model, and was able to account for errors that were heteroscedastic, correlated with the true variables and with other errors, and errors that were non-normally distributed. Expressions of a similar type, however, cannot be obtained for non-linear regression models, such as the logistic regression model.

Correlated errors were investigated by Marshall, Hastrup and Ross[36] for linear or logistic regression analyses with a continuous exposure and confounder measured with error. They found that, in the presence of errors in both explanatory variables, the error correlation inflates or reduces the effect of measurement error on estimated exposure-outcome associations. The bias in the exposure effect estimate may be in either direction when error exists in both explanatory variables and the errors are correlated.

Armstrong, Whittemore and Howe[32] provided formulae describing the effects of measurement error in explanatory variables in multivariable analyses with more than one confounder, assuming a multivariate normal discriminant analysis model for the measurement error. The bias in the exposure effect estimate was shown to have a complex relationship with the covariance matrices of the true explanatory variables and the error variables. Differential measurement errors, such as recall bias in case-control studies, introduced additional bias. For illustration, the effects of measurement error for an exposure and a single confounder on the exposure effect estimate were described. Measurement error in the confounder results in residual confounding of the exposure effect estimate, with the direction of bias dependent on the correlation between exposure and confounder and the relationship of the confounder with outcome. Correlation between the errors in the exposure and confounder reduces the attenuation of the exposure effect estimate due to error in the exposure variable but introduces confounding by the errors. This confounding by errors acts in the same way as residual confounding, but does not necessarily bias effect estimates in the same direction.

2.4.4. Summary

The effects of measurement error or misclassification of confounders on exposure effect estimates are complex, and are summarised below.

- Measurement error in confounders can bias exposure effect estimates towards or away from the null value, even when the error is non-differential and random. For non-differential and random measurement error, and uncorrelated confounders, the observed exposure effect estimate is between the crude estimate and the true effect.
- The observed exposure effect estimates may be biased away from the true value in the opposite direction to the crude estimate, or further from the true value than the crude estimate if the confounder measurement error is systematic.
- Confounder measurement error may cause the observed exposure effect estimate to be in the opposite direction to the true effect if the explanatory variables are correlated.
- Confounder measurement error may induce an apparent exposure effect where there is none.
- Heterogeneity in stratum-specific exposure effect estimates may be caused or masked by confounder measurement error.
- Residual confounding caused by dichotomising a continuous confounder increases with the correlation between exposure and confounder, and is minimised if the cutpoint is the mean of the continuous confounder.
- Confounder measurement error can result in p-values for estimated exposure effects that are too large or too small.
- Correlation between errors in the exposure and confounders reduces the attenuation of the exposure effect estimate due to exposure measurement error but introduces confounding by errors, which may bias the estimate in either direction.

2.5. The impact of unmeasured confounders

In this section, the impact of unmeasured confounding on estimated exposure-outcome effects is described. This concept of the effect of unmeasured confounding was first introduced by Spearman[31] in his 1904 paper. He showed that an unmeasured confounder may either reduce or enlarge the estimated correlation between two variables.

Snedecor and Cochran[42] described the effects of omitting a variable in a linear regression model, where the included variables were assumed to be measured without error. The correct linear regression model is assumed to be

$$Y = \alpha + \beta_E E + \beta_1 X_1 + \varepsilon,$$

where Y is the outcome variable, α is the intercept, β_i are the regression coefficients, E is the exposure, X_1 is the confounder, and ε is the error term with mean zero and uncorrelated with E and X_1. They showed that the estimated exposure-outcome effect, $\hat{\beta}_E$, obtained when the confounder is omitted from the analysis is

Equation 2.2: $$\hat{\beta}_E = \beta_E + \beta_1 b_{1E},$$

where b_{1E} is the coefficient of the linear regression of X_1 on E. If the true model is instead

$$Y = \alpha + \beta_E E + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$

where X_1 and X_2 are both confounders, the estimated exposure-outcome effect when X_2 is omitted from the analysis is

$$\hat{\beta}_E = \beta_E + \beta_2 b_{2E|1},$$

where $b_{2E|1}$ is the regression coefficient for the effect of E on X_2 adjusting for X_1 obtained from a linear regression.

Armstrong[33] provided an expression for the estimated exposure-outcome relative risk (RR), when the RR is an exponential function of the exposure and confounder, the confounder is omitted from the analysis, and where the exposure may be measured with error as Z_E. The estimated exposure-outcome RR was given by

Equation 2.3: $$\hat{\beta}_E = R_{E|1}\beta_E + \frac{\sigma_{Z_E,X_1}}{\sigma^2_{Z_E}}\beta_1$$

where $R_{E|1}$ is the conditional reliability of the exposure given the confounder, σ_{Z_E,X_1} is the covariance between Z_E and X_1, and $\sigma^2_{Z_E}$ is the variance of the error-prone exposure measurements. If E is measured without error, Equation 2.3 is the same as Equation 2.2 provided by Snedecor and Cochran[42] for the effects of an unmeasured confounder in linear regression.

These expressions show that the estimated exposure-outcome effect may be either larger or

smaller than the true exposure-outcome effect in the presence of an unmeasured confounder. The direction in which the estimated effect is biased depends on the true effects of both the exposure and confounder on the outcome, and the covariance between the exposure (either perfectly measured or measured with error) and confounder.

Chapter 3.

The impact of measurement error in explanatory variables and unmeasured confounding in epidemiological studies: A simulation study

3.1. Introduction

In this chapter, the effects of residual and unmeasured confounding and exposure measurement error on estimated exposure-outcome associations are examined using simulation studies. Each simulated dataset contains either two or four confounders, each of which may be affected by measurement error or omitted from the analysis. The exposure may also be measured with error. Measurement errors in exposure and confounders is quantified by the intra-class correlation coefficient (ICC), and ICCs equal to 0.75 and 0.5 are used to generate the mismeasured variables. Logistic regression is used to estimate the exposure-outcome odds ratio (OR) in situations in which the exposure has no true association with outcome (exposure-outcome OR = 1), or in situations in which there is a true association between exposure and outcome (exposure-outcome OR = 2). The results in Section 3.3.1 are the same as those presented by Fewell, Davey Smith and Sterne.[43]

3.2. Methods

3.2.1. Notation and terminology

The following is a list of the notation that will be used throughout this chapter.

E = perfectly measured exposure variable.

Z_E = exposure variable measured with error.

X_i = perfectly measured confounders, for i=1, ..., n.

Z_i = confounders measured with error, for i=1, ..., p.

Y = dichotomous outcome variable.

Throughout this chapter the term *residual confounding* will be used to refer to confounding due to measurement error in a confounder included in a model, and *unmeasured confounding* will be used to refer to confounding due to omission of a confounder from the model.

3.2.2. Dataset structure

Figure 3.1 shows a common situation in epidemiological studies. The dotted arrows indicate possible correlations between exposure and confounders. Often several factors are measured as surrogates for a causal factor, or a causal factor is measured directly but with error. The measured factors Z_E and Z_i, i=1, ..., p, may also be correlated. The measured factors are used to estimate the effect of the causal factors on the dichotomous outcome variable Y. The focus throughout this chapter will be on the simpler case where n=p and each causal factor is represented by one measured factor (Figure 3.2). The causal factors, and therefore the measured factors, can be correlated, as indicated by the dotted arrows.

The logistic model is assumed to relate the causal factors and the outcome.

Equation 3.1: $$\ln\left(\frac{\pi}{1-\pi}\right)=\alpha+\beta_E E+\beta_1 X_1+\ldots+\beta_n X_n$$

Here, π is the probability of the outcome. The exposure and confounders are all normally distributed with mean zero and variance one. Two true exposure-outcome log ORs are considered. In the first, the exposure is assumed to have no causal relationship with the outcome and β_E is 0. The literature described in Chapter 2 shows that the effect of random, non-differential measurement error in the exposure attenuates exposure effect estimates towards the null. In the first situation, the true effect estimate and the null value are equal. For the second scenario, exposure will be assumed to have a causal relationship with outcome, and β_E is ln2 per standard deviation increase. The true confounder-outcome log ORs, β_1, ..., β_n are ln2 per standard deviation increase. The parameter α is set to ln0.1 throughout. Other than changing the number of outcome events in the dataset, the choice of α is not important and changing α would not affect the results presented in this chapter.

Figure 3.1: General relationship between causal factors, measured factors and outcome in an epidemiological study. In the general case, the number of causal factors may differ from the number of measured factors. The measured factors are less than perfect substitutes for the causal factors.

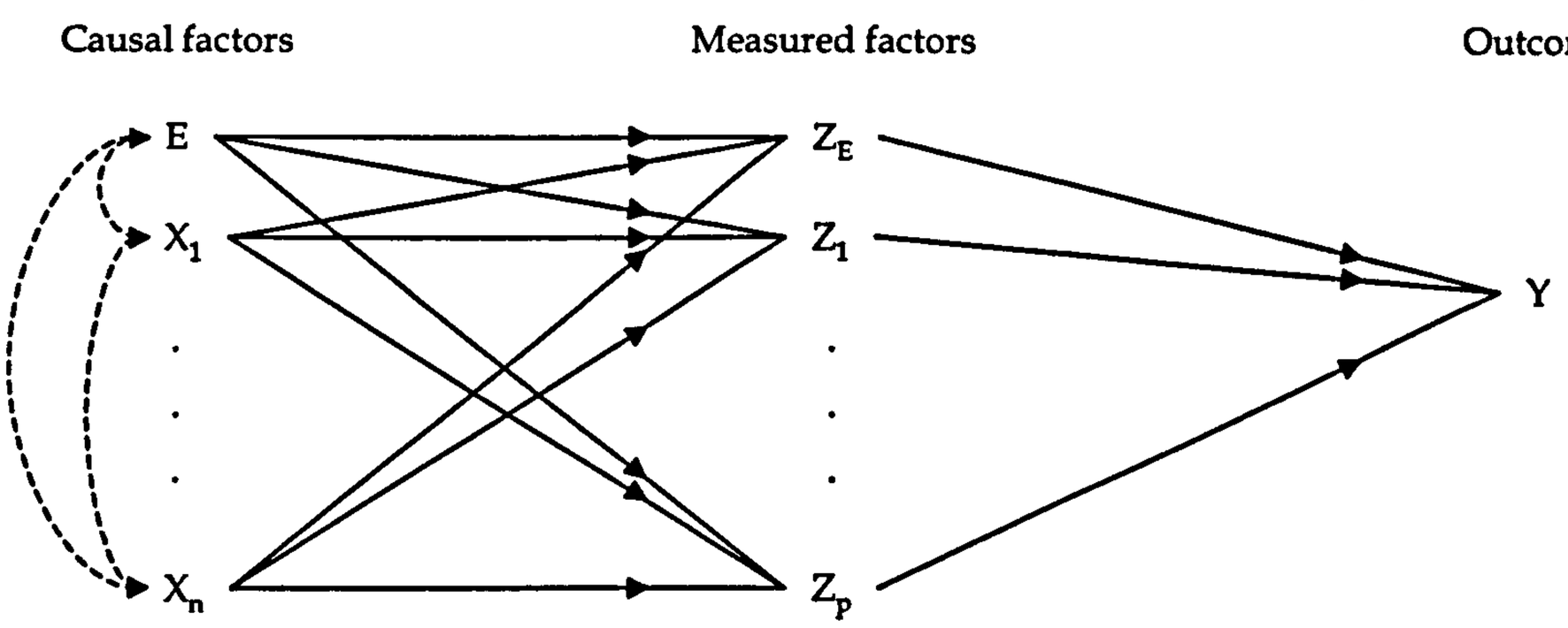
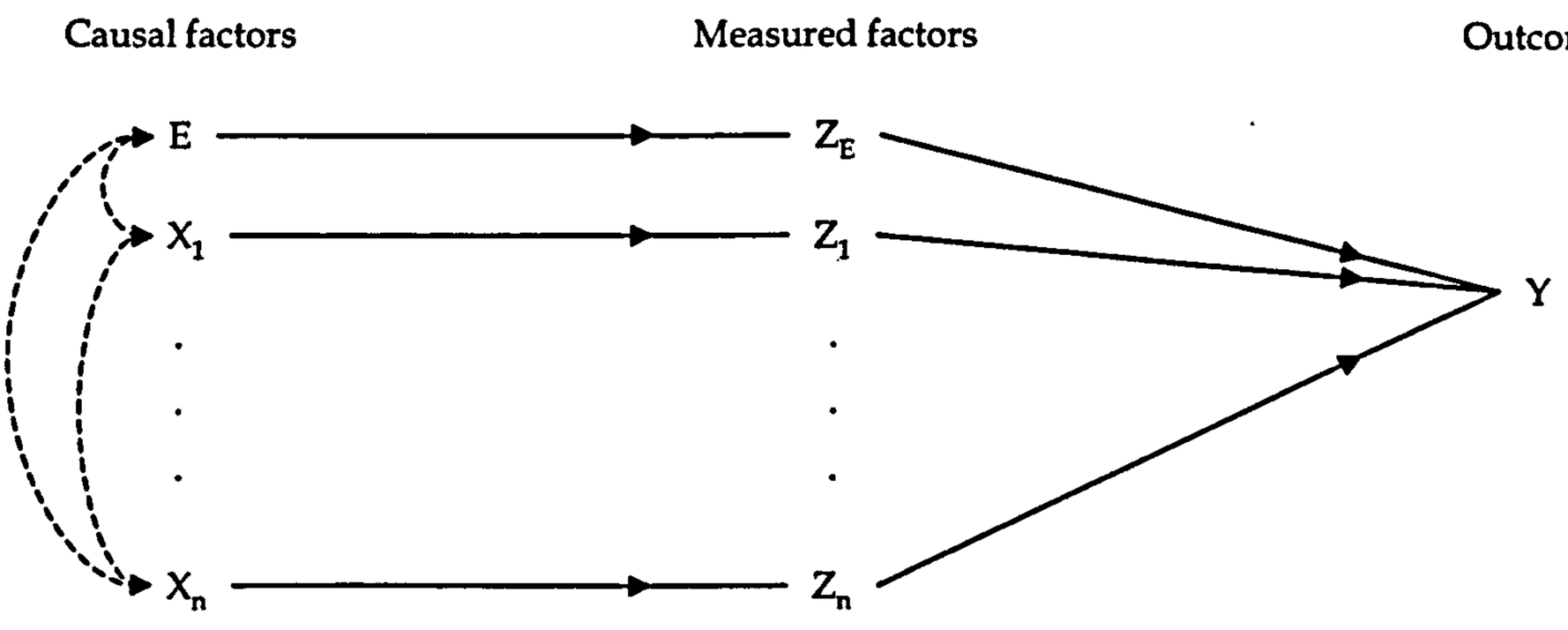


Figure 3.2: Simplified relationship between causal factors, measured factors and outcome in an epidemiological study, assumed in this study. The measured factors are less than perfect substitutes for the causal factors.



Measurement error can be quantified using the ICC, defined as:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

where σ_u^2 is the variance of the true measurements and σ_e^2 is the error variance. An ICC of one implies that $\sigma_e^2=0$ and that there is no measurement error. If $\sigma_e^2=\sigma_u^2$, the ICC will equal 0.5. An ICC of zero implies that $\sigma_u^2=0$, and that the observed variation in the variable is entirely due to error. In epidemiological studies with paired measurements, the usual method for estimating the ICC is to calculate the Pearson correlation with each pair entered twice, once in reverse order. Alternatively, the ICC can be estimated using one-way analysis of variance, or by using a simple random effects model.[44] Error is introduced into the exposure, E, to create the variable Z_E, and into the confounders, $X_1, \ldots, X_n$, to create variables $Z_1, \ldots, Z_n$.

Equation 3.2

$$Z_E = E + \varepsilon_0$$

$$Z_i = X_i + \varepsilon_i \quad \text{for } i=1, \ldots, n$$

where $corr(\varepsilon_i, \varepsilon_j)=0$ if $i \neq j$ for $i,j=0, \ldots, n$ (i.e. measurement errors are independent of each other). For simplicity, it is also assumed that $\varepsilon_i \sim N(0, \sigma_e^2)$, and different distributions for the errors are not considered. The error variance, σ_e^2, is either 1 or 1/3, corresponding to ICCs of 0.5 and 0.75 respectively. These are plausible values for ICCs that can occur in epidemiological studies. For example, Satia-Abouta, Patterson, King *et al.*[45] show a test-retest ICC of 0.74 for iron, and of 0.76 for chromium for mean supplement intake over 10 years from a self-administered questionnaire. Schroder, Covas, Marrugat *et al.*[46] show an ICC of 0.49 for percentage of energy intake in the diet made up by carbohydrate when measured by 72 hour recall, and an ICC of 0.52 for percentage of energy intake in the diet made up by fat when using a food frequency questionnaire. Friesema, Veenstra, Zwietering *et al.*[47] show a test-retest ICC of 0.75 for frequency of drinking in adulthood among women, and of 0.50 for quantity of drinking at age 61 or over among women from the Lifetime Drinking History questionnaire. There are many other examples of ICCs of these sizes in the literature.[48-53] Once the mismeasured exposure variable, Z_E, is created, it is divided by its standard deviation. This results in no change if the ICC of Z_E is one, division by $2\sqrt{3}/3$ if the ICC of Z_E equals 0.75, and division by $\sqrt{2}$ if the ICC of Z_E is 0.5. Using this transformation, estimated ORs for the exposure-outcome association can be interpreted as the OR per standard deviation increase in Z_E.

The first stage in generating the simulated datasets was to draw the exposure E, confounders X_i and error variables ε_i from a multivariate normal distribution, where the correlation between each pair of variables was specified and 500,000 observations per dataset were generated. The mismeasured exposure and confounders were then created according to Equation 3.2. To generate the dichotomous outcome variable, first a uniformly distributed variable between zero and one was generated for each observation in the dataset. The probability of outcome, π, was then calculated for each observation using Equation 3.1 and the values of the parameters

specified previously. From the properties of the uniform distribution, the probability of the uniform variable being less than π equals π. The outcome was therefore defined to be one if the uniform variable was less than π, and zero otherwise. This simulates a Bernoulli trial for each observation in the dataset, and creates an outcome variable that has the correct probability, π, of being equal to one.

The correlations between the exposure E and the confounders were assumed to be 0.1, 0.3 or 0.5, while the correlation between confounders was either 0 or 0.5. Such correlations are not uncommon in epidemiological studies. For example, Osganian, Stampfer, Spiegelman *et al.*[54] report a correlation between serum levels of folic acid and vitamin B_6 of 0.48 and between serum homocysteine and body mass index of 0.09 in a study of 3,524 children from the Child and Adolescent Trial for Cardiovascular Health. They also report that there was no correlation between serum homocysteine and diastolic blood pressure or serum lipids. Variables from the British Women's Heart and Health Study also show correlations of a similar size to those considered here. For example, the correlation between total serum protein concentration and diastolic blood pressure was 0.10, between weight at age 21 and present weight was 0.30, and between serum albumin concentration and mean cellular haemoglobin concentration was 0.49. No generality is lost by considering only positive correlations between exposure and confounders; the corresponding results for negative correlations can be obtained by inverting ORs. Tables of results for correlations between confounders of 0.1, 0.2, 0.3 or 0.4 for the results presented in Section 3.3.1 are provided in Appendix 1.

Once the datasets had been simulated, logistic regression was used to estimate the exposure-outcome OR when the exposure was either perfectly measured or measured with error, and the confounders were either perfectly measured, measured with error, or omitted from the analysis. Each dataset used in the analyses was simulated 50 times, using a different random number seed on each occasion. Odds ratios presented in this chapter are the geometric means of the ORs from these 50 simulations. Fifty simulations each of 500,000 observations resulted in small 95% simulation intervals around the geometric means, with widths no more than 0.011. Stata 8.2 was used for all analyses, using the `corr2data` command with the seed option to generate the normally distributed variables. Two situations are considered in which there are either two confounders or four confounders.

3.3. Residual and unmeasured confounding

In this section, simulations in which the confounders may be measured with error or omitted from the analysis are investigated. The extension to include the situation in which the exposure is also measured with error is provided in Section 3.4.

3.3.1. Exposure unrelated to outcome

First, simulations in which there is no causal effect of exposure on outcome are considered.

3.3.1.1 Two confounders

In this section simulations in which there are two confounding variables, X_1 and X_2, are considered.

Table 3.1 displays estimated ORs for the association between E and the binary outcome, based on simulations in which the confounders are uncorrelated. For each combination of correlations between E and X_1, and E and X_2, the crude OR, the ORs adjusted for Z_1 alone, and the ORs adjusted for both Z_1 and Z_2 are shown. Note that when the ICCs of both Z_1 and Z_2 equal one, then $Z_1=X_1$ and $Z_2=X_2$, so the adjusted OR for E is equal to one (no residual confounding). In all other situations, unmeasured and/or residual confounding bias the estimated OR away from 1. The crude ORs increase with increasing correlations of E with X_1 and X_2, and are symmetric with respect to these correlations. The maximum crude OR is 1.93 when the correlations of E with X_1 and X_2 are each 0.5.

There are nine sets of nine ORs adjusted for both Z_1 and Z_2; these correspond to the residual confounding caused by imperfect measurement of the two confounders. When the correlation between E and X_1 is equal to the correlation between E and X_2, the ORs adjusted for Z_1 and Z_2 are symmetric with respect to the measurement error in Z_1 and Z_2. Where the correlation between E and X_1 is not equal to the correlation between E and X_2, the ORs adjusted for Z_1 and Z_2 are asymmetric with respect to the measurement error. For example, when the correlation of X_1 with E is 0.5, and the correlation of X_2 with E is 0.1, the OR adjusted for both Z_1 and Z_2 is 1.05 when Z_1 is measured with ICC=1 and Z_2 is measured with ICC=0.5. When the ICC of Z_1 is 0.5 and the ICC of Z_2 is 1, the OR adjusted for both Z_1 and Z_2 is 1.22.

When the correlations between E and X_1 and E and X_2 are equal, the OR adjusted for Z_1 only, when Z_1 is measured without error, is equal to the estimated OR adjusted for Z_1 and Z_2 when both have an ICC of 0.5. As the confounders have equal correlation with exposure, their effects on the estimated exposure-outcome OR are equal. An ICC of 0.5 means that half of the variance of the mismeasured confounder is due to error, and therefore controlling for it in the analysis only removes half of its influence on the exposure-outcome OR. Therefore, controlling for two confounders with an ICC of 0.5 removes as much bias from the crude OR as controlling for one confounder with an ICC of 1.

In general, the ORs adjusted for both Z_1 and Z_2 are larger when there is more measurement error (smaller ICC) and when the correlations between E and the confounders are higher. In addition, the ORs adjusted for Z_1 alone are larger than the ORs adjusted for both Z_1 and Z_2

Table 3.1: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.13	1.10	1.07	1.05	1.04	1.29	1.25	1.15	1.10	1.04	1.46	1.43	1.26	1.16	1.05
	0.75		1.09	1.05	1.04	1.02		1.24	1.13	1.08	1.02		1.42	1.24	1.14	1.02
	1		1.07	1.04	1.02	1.00		1.22	1.11	1.06	1.00		1.40	1.22	1.11	1.00
0.3	0.5	1.29	1.18	1.15	1.13	1.11	1.46	1.36	1.24	1.18	1.12	1.67	1.56	1.38	1.27	1.16
	0.75		1.13	1.10	1.08	1.06		1.30	1.18	1.13	1.06		1.50	1.31	1.20	1.08
	1		1.08	1.04	1.02	1.00		1.24	1.12	1.06	1.00		1.44	1.24	1.13	1.00
0.5	0.5	1.46	1.29	1.26	1.24	1.22	1.67	1.51	1.38	1.31	1.24	1.93	1.76	1.56	1.44	1.31
	0.75		1.20	1.16	1.14	1.11		1.41	1.27	1.20	1.13		1.67	1.44	1.31	1.17
	1		1.09	1.05	1.02	1.00		1.30	1.16	1.08	1.00		1.56	1.31	1.17	1.00

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0056.*

because of unmeasured confounding.

Having considered the situation in which the two confounders are uncorrelated, the implications of a correlation between the confounders of 0.5 are now examined. Table 3.2 shows the crude OR, the ORs adjusted for Z_1 alone, and the ORs adjusted for both Z_1 and Z_2 for each combination of correlations of E with X_1 and X_2. Because of the correlation between the confounders the ORs displayed in Table 3.2 are generally smaller than those in Table 3.1. The maximum crude OR is 1.85, when the correlations of E with X_1 and X_2 are 0.5.

In general, as the correlation of E with X_1 and X_2 increases, the OR adjusted for Z_1 alone or both Z_1 and Z_2 increases. There are exceptions to this general rule. For example, when the correlation between E and X_1 is 0.1, the correlation between E and X_2 is 0.5 and Z_1 is measured without error (ICC=1), the OR adjusted for Z_1 only is 1.36 (crude OR 1.43). This is greater than the corresponding OR of 1.25, when the correlation between E and X_1 is 0.5, even though the corresponding crude OR is 1.85. This effect occurs because X_1 is perfectly measured (ICC=1), and is also correlated with X_2. The increased confounding when the correlation between E and X_1 is 0.5 is offset by the correspondingly improved indirect control for the unmeasured confounder (X_2).

Generally, more measurement error in Z_1 or Z_2 (smaller ICC) results in larger estimated ORs (more residual confounding). Again there are exceptions to this general rule. For example, when the correlation between E and X_1 is 0.1, between E and X_2 is 0.5, Z_1 is measured without error (ICC=1) and Z_2 is measured with ICC=0.5, the OR adjusted for both Z_1 and Z_2 is 1.22. When the ICC of Z_1 is 0.5 (measurement error increases), the OR adjusted for Z_1 and Z_2 decreases to 1.20. This effect is due to the partial correlation between the exposure and confounder, controlling for the other confounder.

The partial correlation between E and X_1 controlling for X_2 is defined by:

$$r_{EX_1|X_2} = \frac{r_{EX_1} - r_{EX_2} r_{X_1X_2}}{\sqrt{\left(1 - r_{EX_2}^2\right)\left(1 - r_{X_1X_2}^2\right)}}$$

where r_{EX_1} is the correlation between E and X_1, r_{EX_2} is the correlation between E and X_2, and $r_{X_1X_2}$ is the correlation between the confounders X_1 and X_2. Table 3.3 shows the partial correlation between E and X_1 when controlling for X_2 when the correlation between the confounders is 0.5. With this correlation between confounders, a correlation between E and X_1 of 0.1, and between E and X_2 of 0.5, the partial correlation between E and X_1, conditional on X_2, is negative. In this situation the effect of increasing X_1, while keeping X_2 fixed, is to reduce the estimated exposure-outcome OR. It follows that as measurement error in Z_1 increases, the estimated OR decreases.

Table 3.2: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.13	1.08	1.05	1.04	1.02	1.27	1.23	1.12	1.05	0.98	1.43	1.39	1.20	1.07	0.93
	0.75		1.06	1.04	1.03	1.01		1.21	1.11	1.05	0.99		1.38	1.21	1.10	0.96
	1		1.03	1.02	1.01	1.00		1.18	1.11	1.06	1.00		1.36	1.22	1.13	1.00
0.3	0.5	1.27	1.12	1.12	1.11	1.11	1.43	1.28	1.18	1.13	1.07	1.62	1.46	1.27	1.16	1.03
	0.75		1.04	1.05	1.05	1.06		1.20	1.13	1.09	1.04		1.38	1.23	1.13	1.02
	1		0.97	0.98	0.99	1.00		1.12	1.07	1.04	1.00		1.29	1.18	1.10	1.00
0.5	0.5	1.43	1.17	1.20	1.21	1.22	1.62	1.36	1.27	1.23	1.18	1.85	1.57	1.39	1.27	1.15
	0.75		1.03	1.07	1.10	1.13		1.21	1.16	1.13	1.10		1.42	1.27	1.18	1.08
	1		0.88	0.93	0.96	1.00		1.05	1.03	1.02	1.00		1.25	1.15	1.08	1.00

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0051.*

Table 3.3: Partial correlation between E and X_1, controlling for X_2, by the correlation between E and X_1 and between E and X_2. The correlation between confounders is 0.5.

Correlation between E and X_1	Correlation between E and X_2		
	0.1	0.3	0.5
0.1	0.06	-0.06	-0.20
0.3	0.29	0.18	0.07
0.5	0.52	0.42	0.33

When the correlations between E and X_1 and E and X_2 are equal, controlling for Z_1 only with an ICC of 1 no longer produces the same estimated exposure-outcome OR as controlling for Z_1 and Z_2 when they both have an ICC of 0.5. The correlation between the confounders amplifies the effect of the measurement error, which means that the bias when controlling for both Z_1 and Z_2 with an ICC of 0.5 is worse than when controlling for a perfectly measured Z_2.

Although in general unmeasured confounding results in larger estimated ORs, there are examples where the OR is smaller with greater unmeasured confounding. When the correlation between E and X_1 is 0.5, the correlation between E and X_2 is 0.1, and both confounders have ICC=0.5, the OR adjusted for Z_1 and Z_2 is 1.20. When adjusting only for Z_1, the OR decreases to 1.17. Again, this combination of correlations leads to a negative partial correlation between E and X_2, conditional on X_1. Omitting X_2 from the analysis therefore decreases the estimated exposure-outcome OR.

Intuitively, it would be expected that either unmeasured or residual confounding will lead to imperfect control and hence to adjusted ORs that are intermediate between the crude OR and the correct value of 1.0. However, in Table 3.2, estimated ORs less than one occur when the correlation of E with one of the confounders is 0.1, and the correlation of E with the other confounder is either 0.3 or 0.5. This effect occurs when one confounder is weak (i.e. has a correlation of 0.1 with exposure), and the other is both stronger (correlation of 0.3 or 0.5 with exposure) and is measured without error. Again, this is due to a negative partial correlation between E and the weaker confounder.

3.3.1.2 Four confounders

The results of simulations including four confounders are now considered. To reduce the number of datasets to be simulated, all correlations between pairs of confounders were assumed equal, and were again either 0 or 0.5. The exposure-outcome associations controlling for either one or two confounders were estimated in the analyses.

Table 3.4 displays estimated ORs for the association between exposure and the binary outcome, based on simulations in which the confounders are uncorrelated. For each combination of correlations between E and X_1, and E and X_2, the crude OR, the OR adjusted for Z_1 only, and the OR adjusted for Z_1 and Z_2 are shown. To ensure that the correlation matrix was positive definite, datasets where the correlations between all confounders and exposure was 0.5 were

Table 3.4: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.79	1.76	1.73	1.72	1.71	2.04	2.01	1.91	1.85	1.79	2.35	2.33	2.20	2.13	2.05
	0.75		1.75	1.72	1.71	1.69		2.00	1.89	1.84	1.78		2.32	2.19	2.11	2.03
	1		1.73	1.71	1.69	1.68		1.99	1.88	1.82	1.76		2.31	2.17	2.09	2.00
0.3	0.5	2.04	1.93	1.91	1.89	1.88	2.35	2.24	2.12	2.06	2.00	2.74	2.63	2.52	2.45	2.37
	0.75		1.88	1.85	1.84	1.82		2.18	2.06	2.00	1.93		2.57	2.44	2.37	2.28
	1		1.82	1.79	1.78	1.76		2.12	2.00	1.93	1.87		2.51	2.37	2.29	2.19
0.5	0.5	2.35	2.23	2.20	2.19	2.17	2.74	2.65	2.52	2.44	2.37	3.27	3.22	3.14	3.09	3.03
	0.75		2.16	2.13	2.11	2.09		2.59	2.45	2.37	2.29		3.18	3.09	3.03	2.96
	1		2.09	2.05	2.03	2.00		2.53	2.37	2.28	2.19		3.14	3.03	2.96	2.88

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.011.*

Table 3.5: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.55	1.54	1.53	1.53	1.53	1.71	1.71	1.56	1.47	1.37	1.89	1.92	1.63	1.44	1.22
	0.75		1.54	1.53	1.53	1.53		1.72	1.59	1.50	1.40		1.94	1.69	1.52	1.30
	1		1.53	1.53	1.53	1.54		1.74	1.62	1.54	1.45		1.98	1.77	1.62	1.42
0.3	0.5	1.71	1.53	1.56	1.59	1.62	1.89	1.70	1.58	1.51	1.42	2.10	1.92	1.64	1.47	1.27
	0.75		1.42	1.47	1.50	1.54		1.60	1.51	1.45	1.38		1.81	1.59	1.44	1.26
	1		1.31	1.37	1.40	1.45		1.49	1.42	1.38	1.33		1.70	1.52	1.40	1.25
0.5	0.5	1.89	1.54	1.63	1.69	1.77	2.10	1.74	1.64	1.59	1.52	2.36	1.98	1.71	1.54	1.35
	0.75		1.34	1.44	1.52	1.62		1.53	1.47	1.44	1.40		1.75	1.54	1.41	1.25
	1		1.12	1.22	1.30	1.42		1.29	1.27	1.26	1.25		1.51	1.35	1.25	1.13

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, β_1=β_2=β_3=β_4=ln2. Width of 95% simulation intervals<0.0046.*

not simulated. Table 3.4 therefore displays results where the correlation between E and X_3 is 0.5, and the correlation between E and X_4 is 0.3, as these were the highest correlations available that held for all correlations of E with X_1 and X_2.

Because there is always unmeasured confounding by X_3 and X_4, the estimated ORs in Table 3.4 never equal the correct value of 1. In general, the ORs adjusted for Z_1 alone, or Z_1 and Z_2 are larger when the correlations of E with X_1 and X_2 are larger. More measurement error (smaller ICC) results in larger estimated ORs, and unmeasured confounding increases the estimated ORs. The estimated ORs are larger than those in either Table 3.1 or Table 3.2 due to the larger amount of unmeasured confounding, and crude ORs as large as 3.27 are seen. Residual confounding is now relatively unimportant compared with unmeasured confounding.

Table 3.5 shows results from simulations in which the correlation between the confounders is 0.5. To enable comparison between Table 3.4 and Table 3.5, the correlations between E and X_3, and between E and X_4 are again set to 0.5 and 0.3 respectively.

Because of the correlation between the confounders, the ORs displayed in Table 3.5 are smaller than those in Table 3.4. However they are larger than those in Table 3.1 and Table 3.2 because of the larger amount of unmeasured confounding. Although the relations of the degree of bias due to residual and unmeasured confounding observed in Table 3.4 still hold in general, there are exceptions. There are instances of increasing measurement error (decreasing ICC) or increasing correlation between exposure and confounders leading to smaller estimated ORs. There are also examples of unmeasured confounding leading to decreases in the estimated OR. As discussed in detail in the context of Table 3.2, the correlations between the underlying confounders X_1 to X_4 lead to complex relationships of the adjusted ORs with the strength of confounding (correlation with E), and measurement error in the confounders. Table 3.6 shows the partial correlation between E and X_1, controlling for X_2, X_3 and X_4, for the correlations between confounders used in Table 3.5.

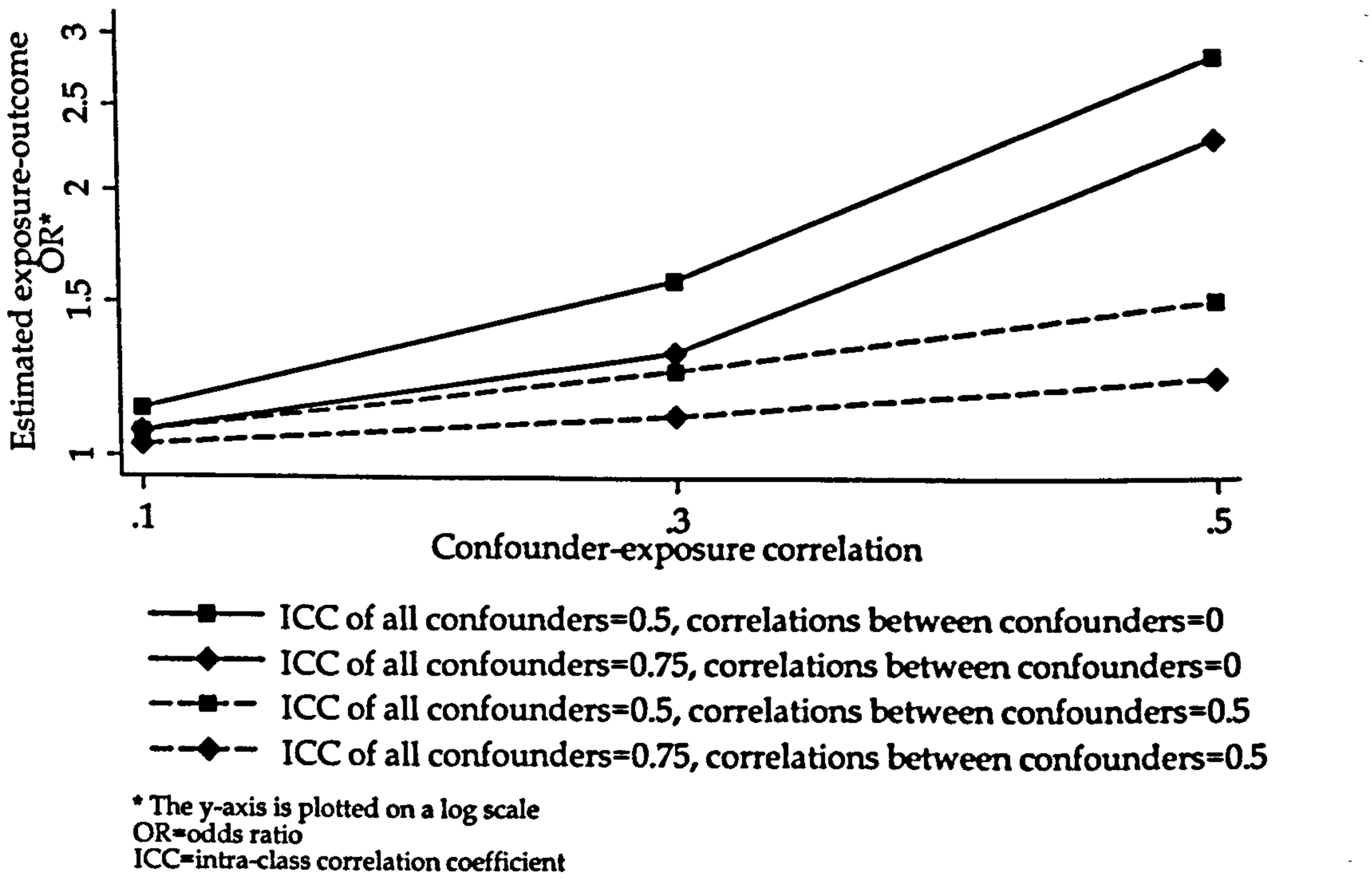
Table 3.6: Partial correlation between E and X_1, controlling for X_2, X_3, and X_4, by the correlation between E and X_1 and between E and X_2. The correlation between confounders is 0.5, between E and X_3 is 0.5, and between E and X_4 is 0.3.

Correlation between E and X_1	Correlation between E and X_2		
	0.1	0.3	0.5
0.1	-0.19	-0.26	-0.35
0.3	0.11	0.04	-0.04
0.5	0.41	0.33	0.27

Figure 3.3 displays the effect of controlling for all four confounders, where the confounders are measured with varying amounts of error. For simplicity, all confounders are assumed to have the same ICC and the correlations between all pairs of confounders are assumed equal, unless the confounder-exposure correlation is 0.5. In this case, the correlation between E and X_4 is 0.3,

and the correlation between all other confounders and exposure is 0.5. The residual confounding of the exposure-outcome OR increases with the measurement error. The adjusted OR increases as the correlation of each confounder with exposure increases, and as the correlations between pairs of confounders decrease. The largest adjusted OR of 2.81 is when the ICC=0.5, the confounders are uncorrelated and the confounder-exposure correlation is 0.5. Figure 3.3 shows that exposure effects may be estimated with substantial bias due to residual confounding alone.

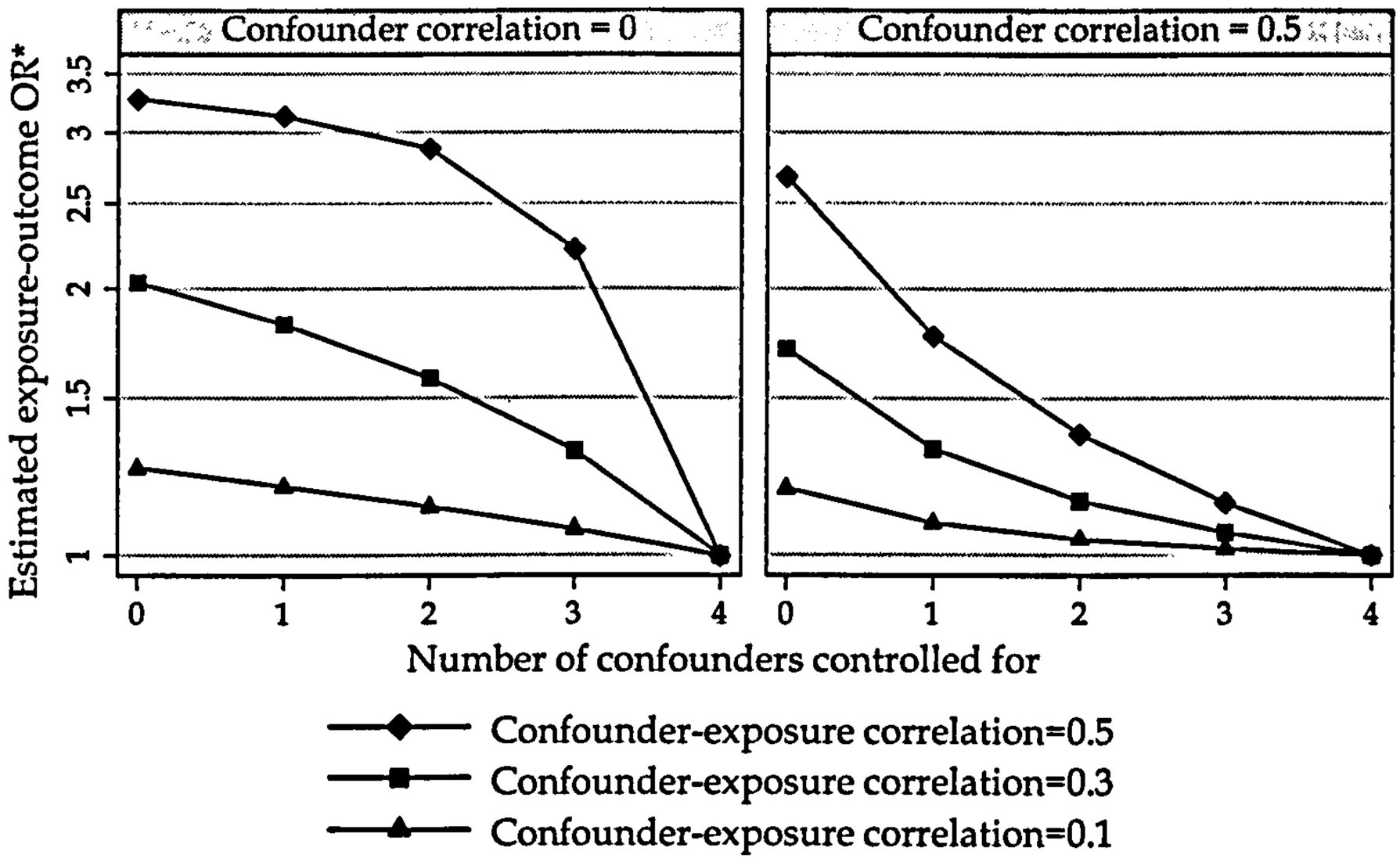
Figure 3.3: Relationship of the exposure-outcome odds ratio adjusted for all four confounders with confounder measurement error, correlation between confounders and confounder-exposure correlation. Measurement error and confounder-exposure correlations are assumed to be equal for all confounders. For a confounder-exposure correlation of 0.5, the correlation between E and X_4 is 0.3.



* The y-axis is plotted on a log scale
OR=odds ratio
ICC=intra-class correlation coefficient

The effect of different numbers of unmeasured confounders on the estimated exposure-outcome OR is displayed in Figure 3.4. For simplicity all confounders are assumed to be measured without error, and the correlations between each confounder and exposure are assumed equal, except when the confounders have a correlation of 0.5, in which case the correlation between E and X_4 is 0.3 while all other confounders have a correlation of 0.5 with exposure. The estimated OR increases as the number of confounders controlled for decreases, and as the correlation between confounders and exposure increases. Bias due to unmeasured confounding is worse when the confounders are uncorrelated. When the exposure-confounder correlation is 0.5, there is serious bias in the estimated exposure-outcome OR of 2.22, even when three confounders are controlled for.

Figure 3.4: Relationship of the exposure-outcome odds ratio with the number of confounders controlled for, confounder-exposure correlation, and correlation between pairs of confounders. All confounders are assumed to be measured without error (ICC=1), and the confounder-exposure correlations are assumed to be equal for all confounders.



* The y-axis is plotted on a log scale
OR=odds ratio
ICC=intra-class correlation coefficient

3.3.2. Exposure related to outcome

In this section, simulations in which the true exposure-outcome OR equals two are considered.

3.3.2.1 Two confounders

Table 3.7 shows the results for simulations with two confounders in which the confounders are uncorrelated. The patterns observed are the same as those observed when there was no true relationship between exposure and outcome in Section 3.3.1.1. In particular, increasing the measurement error in the confounders increases the bias in the estimated exposure-outcome OR, increasing unmeasured confounding by omitting confounders from the analysis increases bias, and increasing the correlation between the exposure and each confounder increases bias. The estimated ORs in Table 3.7 are much larger than those estimated in Table 3.1 due to the increase in the true association between the exposure and outcome.

Table 3.8 displays the results from simulations with two confounders in which the correlation between the confounders equals 0.5. Negative partial correlations between exposure and a confounder, controlling for the other confounder, can lead to reversals of the trends observed when the confounders are uncorrelated. It is possible for increased measurement error in a confounder to decrease the bias in the estimated exposure-outcome OR. Increasing unmeasured confounding by omitting confounders from the analysis can lead to a decrease in

Table 3.7: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	2.09	2.07	2.05	2.04	2.03	2.37	2.36	2.20	2.12	2.03	2.71	2.70	2.40	2.23	2.05
	0.75		2.06	2.04	2.03	2.02		2.35	2.19	2.10	2.02		2.70	2.39	2.22	2.02
	1		2.05	2.03	2.02	2.00		2.34	2.18	2.09	2.00		2.70	2.38	2.20	2.00
0.3	0.5	2.37	2.22	2.20	2.19	2.18	2.71	2.55	2.38	2.29	2.20	3.12	2.96	2.64	2.46	2.26
	0.75		2.14	2.12	2.10	2.09		2.47	2.29	2.20	2.10		2.87	2.54	2.35	2.13
	1		2.06	2.03	2.02	2.00		2.38	2.20	2.10	2.00		2.78	2.43	2.23	2.00
0.5	0.5	2.71	2.43	2.40	2.39	2.38	3.12	2.84	2.64	2.54	2.43	3.65	3.36	3.01	2.81	2.57
	0.75		2.26	2.23	2.22	2.20		2.68	2.46	2.35	2.23		3.20	2.81	2.57	2.31
	1		2.09	2.05	2.02	2.00		2.49	2.26	2.13	2.00		3.01	2.57	2.31	2.00

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=ln2, β_1=β_2=ln2. Width of 95% simulation intervals<0.0099.*

Table 3.8: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	2.00	2.00	2.00	2.00	2.00	2.26	2.27	2.11	2.02	1.92	2.55	2.59	2.26	2.06	1.81
	0.75		2.00	2.00	2.00	2.00		2.28	2.14	2.06	1.96		2.61	2.33	2.14	1.90
	1		2.00	2.00	2.00	2.00		2.29	2.17	2.10	2.00		2.64	2.40	2.23	2.00
0.3	0.5	2.26	2.07	2.11	2.14	2.17	2.55	2.36	2.24	2.17	2.09	2.91	2.72	2.42	2.23	2.01
	0.75		1.97	2.02	2.06	2.10		2.26	2.17	2.11	2.05		2.62	2.36	2.20	2.01
	1		1.86	1.92	1.96	2.00		2.16	2.09	2.05	2.00		2.51	2.31	2.17	2.00
0.5	0.5	2.55	2.16	2.26	2.33	2.40	2.91	2.51	2.42	2.36	2.31	3.36	2.94	2.64	2.45	2.25
	0.75		1.94	2.06	2.14	2.23		2.28	2.23	2.20	2.17		2.69	2.45	2.31	2.14
	1		1.70	1.81	1.90	2.00		2.02	2.01	2.01	2.00		2.42	2.25	2.14	2.00

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=ln2, β_1=β_2=ln2. Width of 95% simulation intervals<0.0086.*

bias, and increasing the correlation between the exposure and each confounder can also lead to a decrease in bias. These effects were all observed in Table 3.2, in which there was no true association between exposure and outcome.

3.3.2.2 Four confounders

Table 3.9 shows the results of simulations in which there are four confounders and the confounders are uncorrelated. These results show that as the correlation between exposure and each confounder increases, so does the bias in the estimated exposure-outcome OR. This was also shown in Table 3.4, where the exposure had no true association with outcome.

In Table 3.9, increasing measurement error in the confounders can cause a decreased bias of the exposure-outcome OR. Consider, for example, when the correlation between E and X_1 equals 0.1 and the correlation between E and X_2 equals 0.5. If only Z_1 is adjusted for in the analysis, and is measured without error, the estimated OR is 4.32. Introducing measurement error into Z_1 so that the ICC equals 0.75 reduces the estimated OR to 4.30. A further increase in the measurement error of Z_1 so that the ICC equals 0.5 results in a further reduction of the estimated OR to 4.28. These are, however, modest differences.

In addition, increased unmeasured confounding can decrease the bias. Consider, for example, when the correlation between E and X_1 is 0.1 and the correlation between E and X_2 is 0.5. If only Z_1 is adjusted for in the analysis, and is measured without error, the estimated exposure-outcome OR is 4.32. When both confounders are omitted from the analysis, the estimated crude exposure-outcome OR is 4.25, which is closer to the true OR of two.

Previously, these counterintuitive effects (i.e. decreasing bias with increasing measurement error or unmeasured confounding) were due to negative partial correlations between the exposure and a confounder. This is not the case in Table 3.9. Instead, these effects are due to the fact that there is now a true association between exposure and outcome. Several authors have shown that the estimated exposure-outcome effect in the presence of residual or unmeasured confounding is a function of the true effects between all explanatory variables and the outcome, and elements of the covariance matrices between the explanatory variables and the errors (see, for example, Armstrong, Whittemore and Howe,[32] or Equation 2.3 derived by Armstrong[33]). Although these expressions have been derived for models other than the logistic regression model used in this chapter, the results are likely to approximately hold. The estimated exposure-outcome OR is therefore composed of two parts; one of which depends on the true exposure-outcome OR, and one of which depends on the true confounder-outcome ORs:

$$\hat{\beta}_E = \alpha_E \beta_E + \sum_{i=1}^{n} \alpha_i \beta_i$$

Table 3.9: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, sample size=500,000, repetitions-50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	3.14	3.13	3.14	3.14	3.14	3.62	3.63	3.47	3.39	3.30	4.25	4.28	4.07	3.94	3.79
	0.75		3.14	3.14	3.14	3.14		3.64	3.48	3.39	3.30		4.30	4.07	3.94	3.79
	1		3.14	3.14	3.14	3.14		3.65	3.48	3.40	3.30		4.32	4.08	3.94	3.78
0.3	0.5	3.62	3.47	3.47	3.48	3.48	4.25	4.09	3.91	3.81	3.72	5.09	4.91	4.71	4.59	4.44
	0.75		3.39	3.39	3.39	3.40		4.00	3.81	3.72	3.62		4.82	4.59	4.46	4.30
	1		3.30	3.30	3.30	3.30		3.91	3.72	3.62	3.51		4.72	4.47	4.32	4.15
0.5	0.5	4.25	4.06	4.07	4.07	4.08	5.09	4.93	4.71	4.59	4.47	6.28	6.17	6.02	5.93	5.82
	0.75		3.94	3.94	3.94	3.94		4.83	4.59	4.46	4.32		6.10	5.93	5.82	5.69
	1		3.81	3.79	3.79	3.78		4.71	4.44	4.30	4.15		6.02	5.82	5.69	5.53

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=ln2, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.019.*

Table 3.10: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	2.40	2.51	2.60	2.67	2.77	2.66	2.81	2.65	2.55	2.44	2.96	3.19	2.77	2.50	2.18
	0.75		2.59	2.67	2.74	2.82		2.93	2.77	2.67	2.56		3.35	2.96	2.70	2.37
	1		2.69	2.77	2.82	2.90		3.08	2.93	2.83	2.72		3.57	3.22	2.97	2.64
0.3	0.5	2.66	2.48	2.65	2.77	2.93	2.96	2.80	2.69	2.62	2.55	3.33	3.18	2.79	2.56	2.27
	0.75		2.39	2.55	2.67	2.83		2.71	2.62	2.57	2.52		3.09	2.77	2.55	2.29
	1		2.28	2.44	2.56	2.72		2.61	2.55	2.52	2.48		3.01	2.74	2.55	2.32
0.5	0.5	2.96	2.51	2.77	2.96	3.22	3.33	2.85	2.79	2.77	2.74	3.79	3.28	2.91	2.68	2.41
	0.75		2.23	2.50	2.70	2.97		2.57	2.56	2.55	2.55		2.97	2.68	2.49	2.27
	1		1.93	2.18	2.37	2.64		2.24	2.27	2.29	2.32		2.63	2.41	2.27	2.10

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=ln2, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0096.*

where the α are constants which depend on elements of the covariance matrix between the true exposure and confounders and elements of the covariance matrix between the errors. As measurement error in the confounders increases, α_E may decrease. It is also possible that the decrease in $\alpha_E\beta_E$, caused by increased measurement error in the confounders, is greater than the corresponding increase in $\sum_{i=1}^{n}\alpha_i\beta_i$. Therefore, reduced bias may be observed with increased measurement error or unmeasured confounding. When there is no true association between exposure and outcome, as in Section 3.3.1, measurement error in the confounders cannot cause a decrease in $\alpha_E\beta_E$, because β_E is equal to zero and therefore patterns of bias are more predictable.

Table 3.10 shows the results of simulations in which there are four confounders, the correlations between pairs of confounders equals 0.5, and the true exposure-outcome OR is two. Bias in the estimated exposure-outcome OR can be decreased with increased measurement error in the confounders, increased unmeasured confounding, and increased correlation between the exposure and each confounder. This is caused by the correlation between the confounders. These effects were observed in Table 3.5, where there was no true association between exposure and outcome.

3.4. Residual and unmeasured confounding and exposure measurement error

In this section, the results presented in Section 3.3 are extended to the situation in which the exposure is also measured with error.

3.4.1. Exposure unrelated to outcome

First, simulations in which there is no causal effect of exposure on outcome are investigated.

3.4.1.1 Two confounders

The results of simulations in which there are two confounders, where the confounders and exposure can be measured with error are now considered.

Table 3.11 and Table 3.12 show the effect of measurement error in the exposure and confounders for exposure ICCs of 0.75 and 0.5 respectively, for the situation in which the confounders are uncorrelated.

Table 3.11 and Table 3.12 show that the estimated OR increases with increasing measurement error in the confounders (decreasing ICC) and with greater correlation between the exposure and confounders. Unmeasured confounding, caused by omitting Z_2 from the analysis, also

Table 3.11: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, ICC of exposure=0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.11	1.09	1.06	1.04	1.03	1.24	1.21	1.12	1.08	1.03	1.39	1.36	1.21	1.12	1.04
	0.75		1.07	1.04	1.03	1.02		1.20	1.11	1.06	1.02		1.35	1.19	1.11	1.02
	1		1.06	1.03	1.02	1.00		1.19	1.09	1.05	1.00		1.33	1.18	1.09	1.00
0.3	0.5	1.24	1.15	1.12	1.11	1.09	1.39	1.29	1.20	1.15	1.10	1.55	1.46	1.30	1.21	1.12
	0.75		1.11	1.08	1.06	1.05		1.25	1.15	1.10	1.05		1.41	1.24	1.15	1.06
	1		1.06	1.03	1.02	1.00		1.20	1.10	1.05	1.00		1.36	1.19	1.10	1.00
0.5	0.5	1.39	1.24	1.21	1.19	1.18	1.55	1.40	1.30	1.24	1.19	1.75	1.59	1.42	1.33	1.22
	0.75		1.16	1.12	1.11	1.09		1.32	1.21	1.15	1.10		1.51	1.33	1.23	1.12
	1		1.07	1.04	1.02	1.00		1.23	1.12	1.06	1.00		1.42	1.22	1.12	1.00

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0042.*

Table 3.12: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.09	1.07	1.05	1.04	1.02	1.19	1.17	1.10	1.06	1.02	1.30	1.28	1.16	1.09	1.03
	0.75		1.06	1.04	1.02	1.01		1.16	1.09	1.05	1.01		1.27	1.15	1.08	1.01
	1		1.05	1.02	1.01	1.00		1.15	1.08	1.04	1.00		1.26	1.14	1.07	1.00
0.3	0.5	1.19	1.12	1.10	1.09	1.08	1.30	1.23	1.16	1.12	1.08	1.43	1.35	1.22	1.16	1.09
	0.75		1.09	1.06	1.05	1.04		1.19	1.12	1.08	1.04		1.31	1.18	1.11	1.04
	1		1.05	1.02	1.01	1.00		1.16	1.08	1.04	1.00		1.27	1.14	1.07	1.00
0.5	0.5	1.30	1.18	1.16	1.15	1.14	1.43	1.30	1.22	1.18	1.14	1.57	1.44	1.30	1.23	1.16
	0.75		1.12	1.09	1.08	1.07		1.24	1.16	1.11	1.07		1.37	1.23	1.16	1.08
	1		1.05	1.03	1.01	1.00		1.17	1.09	1.04	1.00		1.30	1.16	1.08	1.00

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0037.*

increases the estimated OR. These trends are the same as those observed in Table 3.1, in which there was no measurement error in the exposure.

The estimated ORs in Table 3.12 are smaller than those in Table 3.1, and the ORs presented in Table 3.11 are intermediate between the values shown in Table 3.1 and Table 3.12. For example, the maximum crude OR in Table 3.1 is 1.93 when the correlations between both confounders and exposure are 0.5. The corresponding crude OR in Table 3.12 is 1.57, and in Table 3.11 is 1.75. The reasoning behind this effect is straightforward. Non-differential and random measurement error in the exposure attenuates the estimated exposure-outcome OR towards the null value. As the exposure measurement error for the values displayed in Table 3.12 is greater than that in Table 3.11, with an ICC of exposure equal to 0.5 rather than 0.75, the ORs in Table 3.12 show a greater attenuation towards the null value of one.

Note that, although the exposure is measured with error in Table 3.11 and Table 3.12, there is no bias in the estimated exposure-outcome OR if both confounders are included in the analysis and measured perfectly. The estimated OR observed when exposure is measured with error will be closer to the null value than the estimated OR when there is no exposure measurement error. If the estimated OR equals the null value when there is no exposure measurement error, then it follows that introducing measurement error into the exposure will have no effect on the estimated exposure-outcome OR, and will still produce an OR equal to one.

Having considered the effect of measurement error in the exposure and confounders when the confounders are uncorrelated, the scenario in which the confounders are correlated is now investigated.

Table 3.13 and Table 3.14 display the estimated ORs from simulations in which the exposure has an ICC of 0.75 and 0.5 respectively. The confounders are assumed to have a correlation of 0.5. Although in general the estimated ORs increase with increasing measurement error in the confounders, increasing correlation between the confounders and exposure, and increased unmeasured confounding, there are exceptions to all of these generalisations.

For example, bias in the estimated OR decreases as measurement error increases when the correlation between E and X_1 equals 0.1, and between E and X_2 equals 0.5 in both Table 3.13 and Table 3.14. Considering Table 3.14, if the ICC of Z_1 is one and the ICC of Z_2 is 0.5, the estimated OR is 1.14. Increasing measurement error in Z_1 so that the ICC equals 0.5 results in an estimated OR of 1.12, which is closer to the true OR of one.

Reduced bias is observed with increased confounder-exposure correlation when the correlation between E and X_2 equals 0.3, the ICC of Z_1 equals one, and the ICC of Z_2 equals 0.5. When the

Table 3.13: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, ICC of exposure=0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.11	1.07	1.05	1.03	1.02	1.23	1.19	1.10	1.04	0.98	1.36	1.33	1.16	1.06	0.94
	0.75		1.05	1.03	1.02	1.01		1.17	1.09	1.05	0.99		1.31	1.17	1.08	0.97
	1		1.03	1.02	1.01	1.00		1.16	1.09	1.05	1.00		1.30	1.18	1.10	1.00
0.3	0.5	1.23	1.10	1.10	1.09	1.09	1.36	1.23	1.15	1.11	1.06	1.51	1.38	1.22	1.13	1.02
	0.75		1.04	1.04	1.05	1.05		1.17	1.11	1.07	1.03		1.31	1.18	1.10	1.01
	1		0.97	0.98	0.99	1.00		1.10	1.06	1.03	1.00		1.24	1.14	1.08	1.00
0.5	0.5	1.36	1.14	1.16	1.17	1.18	1.51	1.29	1.22	1.18	1.14	1.69	1.45	1.30	1.21	1.11
	0.75		1.02	1.06	1.08	1.10		1.16	1.13	1.10	1.08		1.33	1.21	1.14	1.06
	1		0.90	0.94	0.97	1.00		1.03	1.02	1.01	1.00		1.19	1.11	1.06	1.00

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0039.*

Table 3.14: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.09	1.06	1.04	1.03	1.01	1.18	1.15	1.08	1.03	0.99	1.28	1.26	1.12	1.04	0.95
	0.75		1.04	1.03	1.02	1.01		1.14	1.07	1.04	0.99		1.25	1.13	1.06	0.97
	1		1.02	1.01	1.01	1.00		1.12	1.07	1.04	1.00		1.24	1.14	1.08	1.00
0.3	0.5	1.18	1.08	1.08	1.07	1.07	1.28	1.18	1.12	1.08	1.04	1.40	1.29	1.17	1.10	1.01
	0.75		1.03	1.03	1.04	1.04		1.13	1.08	1.06	1.02		1.24	1.14	1.08	1.01
	1		0.98	0.99	0.99	1.00		1.08	1.04	1.02	1.00		1.19	1.11	1.06	1.00
0.5	0.5	1.28	1.11	1.12	1.13	1.14	1.40	1.22	1.17	1.14	1.11	1.53	1.34	1.22	1.16	1.08
	0.75		1.02	1.04	1.06	1.08		1.12	1.10	1.08	1.06		1.24	1.16	1.10	1.04
	1		0.92	0.95	0.97	1.00		1.03	1.01	1.01	1.00		1.14	1.08	1.04	1.00

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_0=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0034.*

correlation between E and X_1 equals 0.1, the estimated OR in Table 3.14 equals 1.07. Increasing the correlation between E and X_1 to 0.5 decreases the estimated OR to 1.01.

Decreased bias with increased unmeasured confounding is observed when the correlation between E and X_1 equals 0.5 and between E and X_2 equals 0.1. When the ICCs of both confounders are 0.5, and both are included in the analysis, the estimated OR in Table 3.14 is 1.12. When Z_2 is omitted from the analysis, the estimated OR is 1.11.

These reversals of trend are due to the partial correlations between exposure and confounders. These have been shown in Table 3.3.

The estimated ORs in Table 3.13 and Table 3.14 are all less than those in Table 3.2, due to measurement error in the exposure. In addition, the values displayed in Table 3.13 are intermediate between the values estimated when the exposure is measured without error (Table 3.2), and when the exposure ICC equals 0.5 (Table 3.14). For example, the maximum crude OR estimated in Table 3.2 is 1.85 when the correlations between the confounders and exposure are 0.5. The corresponding OR in Table 3.14 is 1.53, and in Table 3.13 is 1.69. Again, measurement error in the exposure variable does not produce any bias in the estimated exposure-outcome OR if both confounders are measured perfectly and included in the analysis.

3.4.1.2 Four confounders

The situation in which there are four confounders is now considered. The exposure and confounders may all be measured with error.

Table 3.15 and Table 3.16 show the results of simulations with four confounders in which the exposure ICC is 0.75 and 0.5 respectively, and the correlation between the perfectly measured confounders is 0. The trends in exposure-outcome OR found in Table 3.4 again hold for Table 3.15 and Table 3.16. The estimated OR increases as the correlation between E and X_1 or X_2 increases, as measurement error in the confounders increases (ICC decreases), and as unmeasured confounding increases. The ORs displayed in Table 3.15 and Table 3.16 are all smaller than those in Table 3.4. Furthermore, the ORs in Table 3.16 are all smaller than those in Table 3.15. This is due to increasing measurement error in the exposure variable attenuating the exposure-outcome effect estimate. For example, the maximum crude OR in Table 3.4 occurs when the correlation between E and X_1, and E and X_2 are both 0.5, and is equal to 3.27. The corresponding crude OR in Table 3.15 is 2.64, and in Table 3.16 is 2.13. The estimated ORs adjusted for the four perfectly measured confounders are not displayed in Table 3.15 and Table 3.16, but again exposure measurement error does not create any bias in this situation.

Having considered the implications of exposure measurement error when the confounders are

Table 3.15: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, ICC of exposure=0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.64	1.62	1.60	1.58	1.57	1.83	1.81	1.72	1.67	1.62	2.05	2.03	1.90	1.83	1.75
	0.75		1.61	1.58	1.57	1.56		1.80	1.71	1.66	1.61		2.02	1.89	1.81	1.73
	1		1.60	1.57	1.56	1.55		1.79	1.69	1.65	1.60		2.01	1.88	1.80	1.72
0.3	0.5	1.83	1.74	1.72	1.71	1.69	2.05	1.95	1.86	1.81	1.76	2.31	2.22	2.08	2.01	1.92
	0.75		1.69	1.67	1.66	1.65		1.91	1.81	1.76	1.71		2.17	2.03	1.95	1.86
	1		1.65	1.62	1.61	1.60		1.86	1.76	1.71	1.65		2.11	1.97	1.89	1.80
0.5	0.5	2.05	1.92	1.90	1.89	1.88	2.31	2.19	2.08	2.03	1.97	2.64	2.52	2.39	2.31	2.22
	0.75		1.85	1.83	1.81	1.80		2.12	2.01	1.95	1.89		2.46	2.31	2.22	2.13
	1		1.77	1.75	1.73	1.72		2.05	1.92	1.86	1.80		2.39	2.22	2.13	2.02

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, β_1=β_2=β_3=β_4=ln2. Width of 95% simulation intervals<0.0062.*

Table 3.16: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.49	1.47	1.46	1.45	1.44	1.62	1.61	1.54	1.50	1.46	1.77	1.76	1.64	1.58	1.51
	0.75		1.46	1.45	1.44	1.43		1.60	1.53	1.49	1.45		1.75	1.63	1.57	1.51
	1		1.46	1.44	1.43	1.42		1.59	1.52	1.48	1.45		1.74	1.62	1.56	1.50
0.3	0.5	1.62	1.55	1.54	1.53	1.52	1.77	1.70	1.62	1.59	1.55	1.94	1.86	1.74	1.68	1.61
	0.75		1.52	1.50	1.49	1.48		1.66	1.59	1.55	1.51		1.83	1.70	1.64	1.57
	1		1.48	1.46	1.45	1.45		1.62	1.55	1.51	1.47		1.79	1.66	1.60	1.53
0.5	0.5	1.77	1.66	1.64	1.63	1.62	1.94	1.82	1.74	1.70	1.66	2.13	2.02	1.89	1.82	1.75
	0.75		1.60	1.58	1.57	1.56		1.76	1.68	1.64	1.60		1.95	1.82	1.75	1.68
	1		1.53	1.51	1.51	1.50		1.70	1.61	1.57	1.53		1.89	1.75	1.68	1.60

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0043.*

Table 3.17: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, ICC of exposure=0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.46	1.45	1.44	1.44	1.44	1.58	1.58	1.46	1.38	1.30	1.72	1.74	1.50	1.35	1.17
	0.75		1.44	1.44	1.44	1.44		1.59	1.48	1.41	1.33		1.76	1.54	1.40	1.23
	1		1.44	1.44	1.44	1.44		1.60	1.50	1.44	1.36		1.78	1.60	1.47	1.31
0.3	0.5	1.58	1.43	1.46	1.48	1.50	1.72	1.57	1.47	1.41	1.34	1.87	1.73	1.50	1.37	1.21
	0.75		1.35	1.38	1.41	1.44		1.49	1.41	1.36	1.31		1.65	1.46	1.34	1.20
	1		1.26	1.30	1.33	1.36		1.40	1.34	1.31	1.27		1.56	1.41	1.31	1.19
0.5	0.5	1.72	1.43	1.50	1.54	1.60	1.87	1.58	1.50	1.46	1.41	2.05	1.75	1.54	1.41	1.27
	0.75		1.27	1.35	1.40	1.47		1.41	1.37	1.34	1.31		1.57	1.41	1.31	1.19
	1		1.09	1.17	1.23	1.31		1.23	1.21	1.20	1.19		1.38	1.27	1.19	1.10

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, β_1=β_2=β_3=β_4=ln2. Width of 95% simulation intervals<0.0041.*

Table 3.18: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, ICC of exposure=0.5, sample size=500,000, repetitions=50.

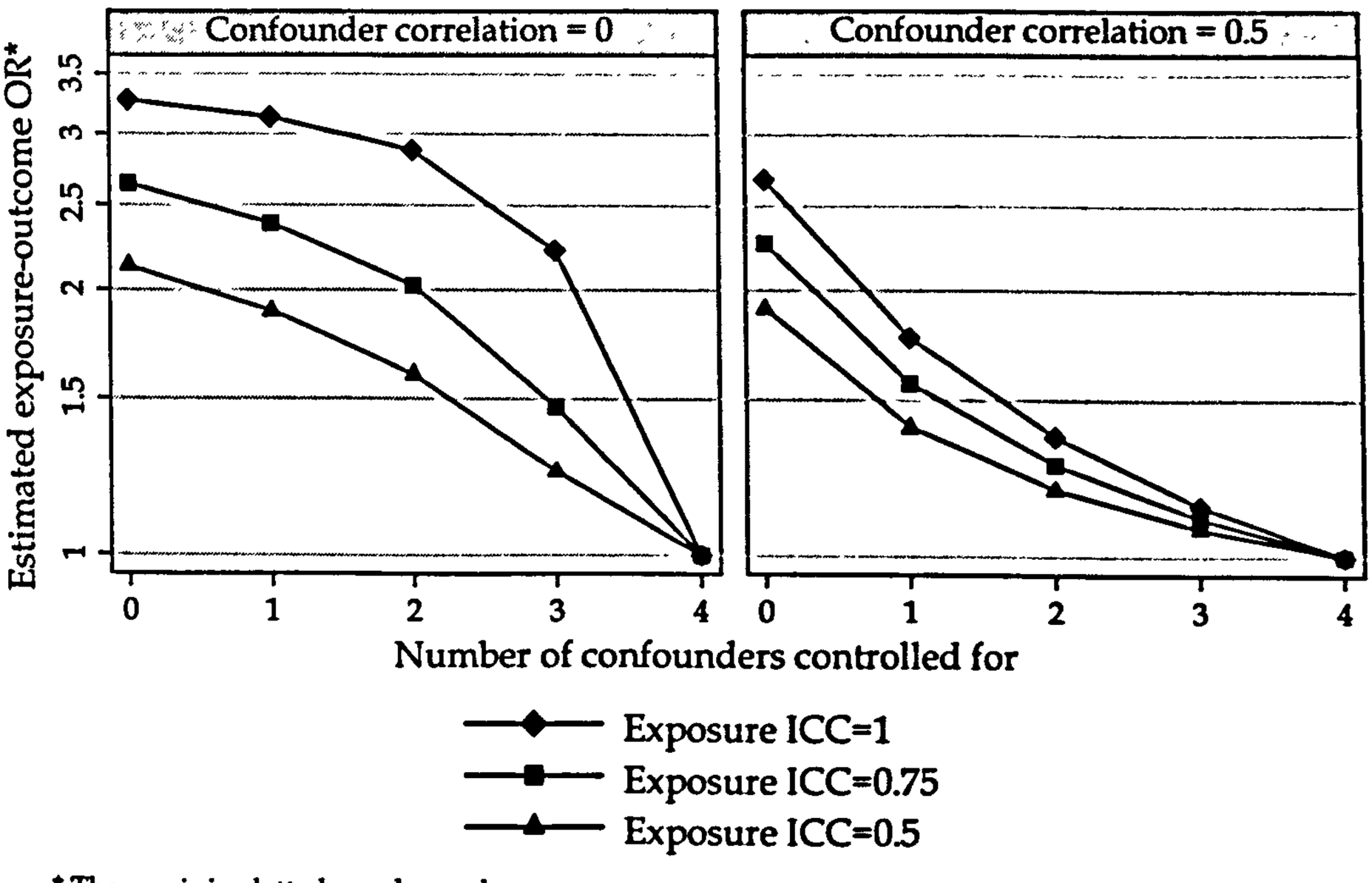
Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.36	1.35	1.34	1.34	1.34	1.45	1.45	1.35	1.29	1.23	1.54	1.56	1.37	1.26	1.13
	0.75		1.34	1.34	1.34	1.34		1.45	1.37	1.31	1.25		1.57	1.40	1.29	1.17
	1		1.34	1.34	1.34	1.34		1.46	1.38	1.34	1.28		1.59	1.44	1.34	1.23
0.3	0.5	1.45	1.33	1.35	1.37	1.38	1.54	1.43	1.36	1.31	1.26	1.65	1.54	1.37	1.27	1.16
	0.75		1.27	1.29	1.31	1.34		1.37	1.31	1.28	1.24		1.48	1.34	1.25	1.15
	1		1.20	1.23	1.25	1.28		1.30	1.26	1.24	1.20		1.42	1.30	1.23	1.14
0.5	0.5	1.54	1.32	1.37	1.40	1.44	1.65	1.43	1.37	1.34	1.30	1.77	1.54	1.39	1.30	1.19
	0.75		1.20	1.26	1.29	1.34		1.30	1.27	1.25	1.23		1.41	1.30	1.22	1.14
	1		1.07	1.13	1.17	1.23		1.17	1.16	1.15	1.14		1.28	1.19	1.14	1.07

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, β_1=β_2=β_3=β_4=ln2. Width of 95% simulation intervals<0.0033.*

uncorrelated, correlations between confounders of 0.5 are now considered. Table 3.17 and Table 3.18 show the estimated exposure-outcome ORs for exposure ICCs of 0.75 and 0.5 respectively in this setting.

Although the general trends in the estimated exposure-outcome OR estimated in Table 3.15 and Table 3.16 still hold in general, there are exceptions to the rules. In Table 3.17 and Table 3.18, the estimated ORs can decrease as the correlation between exposure and the confounders increases, as measurement error in the confounders increases, and as unmeasured confounding increases. This was also shown in Table 3.5 when the exposure was measured without error. Again, the ORs presented in Table 3.17 and Table 3.18 are smaller than those shown in Table 3.5, and the ORs shown in Table 3.18 are smaller than those in Table 3.17, which indicates that increasing exposure measurement error attenuates the estimated OR in this setting. For example, the maximum crude OR in Table 3.5 is 2.36. The corresponding crude OR in Table 3.17 is 2.05, and in Table 3.18 is 1.77. The correlation between the confounders means that the ORs shown in Table 3.18 are smaller than those in Table 3.16, and the ORs in Table 3.17 are smaller than those in Table 3.15.

Figure 3.5: The effect of exposure measurement error on the estimated exposure-outcome OR.



* The y-axis is plotted on a log scale
OR=odds ratio
ICC=intra-class correlation coefficient

Figure 3.5 demonstrates the effect of exposure measurement error on the estimated exposure-outcome OR. When the confounder correlation is 0, the figure displays estimated exposure-outcome ORs when the correlation between exposure and the confounders X_1, X_2 and X_3 are all 0.5, and the correlation between E and X_4 is 0.3. When the correlation between the confounders

is 0.5, the correlation between E and all confounders is 0.5.

As seen previously in Figure 3.4, the effect of unmeasured confounding on the exposure OR is greater when the confounders are uncorrelated. This effect still occurs when the exposure is measured with error, although the difference between corresponding ORs estimated when the confounder correlation is 0 and 0.5 decreases. The attenuation effect of exposure measurement error is greater when the confounders are uncorrelated, which can be seen by the larger distance between the lines.

These results have shown the joint effect of residual and unmeasured confounding and exposure measurement error in simulations in which there is no true relationship between exposure and outcome. In this situation, the bias induced by measurement error in the exposure attenuates the estimated ORs towards the true value, as it happens to equal the null value. It is of interest to investigate the effects of exposure measurement error in situations in which the exposure has a true effect on outcome. These results are now presented.

3.4.2. Exposure related to outcome

3.4.2.1 Two confounders

First the situation in which there are two confounders is considered. When the exposure is measured without error, the same patterns of bias are observed as when the exposure has no true association with the outcome. Tables presenting these results have been described in Section 3.3.2.1.

Table 3.19 and Table 3.20 show the results obtained when the ICC of exposure is 0.75 and 0.5 respectively and the two confounders are uncorrelated. In both tables, the estimated OR is larger when the confounders are measured with more error (smaller ICC). Unmeasured confounding, such as including only Z_1 in the analysis or omitting both confounders, increases the estimated OR. When the ICC of exposure equals 0.75, the estimated ORs are smaller than those estimated when exposure is measured without error (see Section 3.3.2.1). Increasing the measurement error in the exposure so that the ICC is equal to 0.5 results in a further decrease in the estimated ORs. These trends are the same as those observed in Table 3.11 and Table 3.12.

The estimated ORs in Table 3.11 and Table 3.12 were generally larger than the true OR. In Table 3.19 and Table 3.20, the estimated ORs generally underestimate the association between exposure and outcome. Residual and unmeasured confounding, however, mean that it is also possible to overestimate the true exposure-outcome OR.

In contrast with Table 3.11 and Table 3.12, an increase in the correlation between exposure and

Table 3.19: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, ICC of exposure=0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.87	1.85	1.83	1.83	1.82	2.07	2.06	1.93	1.86	1.80	2.30	2.29	2.04	1.90	1.75
	0.75		1.84	1.83	1.82	1.81		2.05	1.92	1.86	1.79		2.29	2.03	1.89	1.74
	1		1.83	1.82	1.81	1.80		2.04	1.91	1.85	1.78		2.28	2.02	1.88	1.72
0.3	0.5	2.07	1.95	1.93	1.92	1.91	2.30	2.17	2.04	1.97	1.90	2.56	2.43	2.17	2.02	1.87
	0.75		1.88	1.86	1.86	1.85		2.11	1.97	1.90	1.83		2.36	2.09	1.94	1.78
	1		1.82	1.80	1.79	1.78		2.04	1.90	1.83	1.75		2.29	2.01	1.86	1.69
0.5	0.5	2.30	2.05	2.04	2.03	2.02	2.56	2.31	2.17	2.09	2.01	2.86	2.61	2.33	2.17	1.99
	0.75		1.92	1.90	1.89	1.88		2.18	2.02	1.94	1.86		2.48	2.17	1.99	1.81
	1		1.78	1.75	1.74	1.72		2.03	1.87	1.78	1.69		2.33	1.99	1.81	1.61

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_x=ln2, β_1=β_2=ln2. Width of 95% simulation intervals<0.0078.*

Table 3.20: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, ICC of exposure = 0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.65	1.64	1.62	1.62	1.61	1.78	1.77	1.68	1.63	1.58	1.93	1.92	1.73	1.63	1.52
	0.75		1.63	1.62	1.61	1.61		1.77	1.67	1.63	1.58		1.92	1.72	1.62	1.51
	1		1.62	1.61	1.61	1.60		1.76	1.67	1.62	1.57		1.91	1.72	1.61	1.50
0.3	0.5	1.78	1.69	1.68	1.67	1.67	1.93	1.84	1.74	1.69	1.64	2.08	1.99	1.79	1.69	1.58
	0.75		1.64	1.63	1.63	1.62		1.79	1.69	1.64	1.59		1.94	1.74	1.63	1.52
	1		1.60	1.58	1.58	1.57		1.74	1.64	1.59	1.53		1.90	1.69	1.58	1.46
0.5	0.5	1.93	1.74	1.73	1.72	1.72	2.08	1.90	1.79	1.74	1.69	2.26	2.06	1.86	1.75	1.63
	0.75		1.64	1.63	1.62	1.61		1.79	1.69	1.63	1.58		1.96	1.75	1.63	1.50
	1		1.54	1.52	1.51	1.50		1.69	1.58	1.52	1.46		1.86	1.63	1.51	1.38

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=ln2, β_1=β_2=ln2. Width of 95% simulation intervals<0.0058.*

confounders does not always result in an increase in the estimated OR. Consider, for example, Table 3.19. When the correlation between E and X_1 is 0.1, the correlation between E and X_2 is 0.5, and both confounders are perfectly measured, the estimated OR is 1.72. Increasing the correlation between E and X_1 to 0.3 leads to a decrease in the estimated OR to 1.69. A further increase in the correlation between E and X_1 to 0.5 creates an estimated OR of 1.61, which is again smaller. This effect is expected as, in the absence of residual and unmeasured confounding, the attenuation effect of exposure measurement error will be amplified by correlations between the exposure and confounders.

More generally, this effect occurs when one of the confounders is perfectly measured (ICC=1), and the correlation between the perfectly measured confounder and exposure increases. In this situation, the exposure error is acting to decrease the OR, and this effect is amplified by correlations between exposure and confounders. There is also the additional effect of residual confounding, which acts to increase the OR, with the effect also being amplified by the correlation between the confounder and exposure. When one confounder is measured without error, the increase of the OR caused by residual confounding is not as large as the reduction in the OR due to exposure error. The overall effect is that the estimated OR decreases as the correlation between exposure and confounder increases.

Non-monotonic trends of the estimated OR with the correlation between exposure and confounders can be seen in Table 3.20. For example, when the correlations between both confounders and exposure are equal to 0.1, the ICC of Z_1 equals 0.75 and Z_2 is measured without error, the estimated OR is 1.61. When the correlation between E and X_1 increases to 0.3, the estimated OR increases to 1.62. When the correlation between E and X_1 increases further to 0.5, the estimated OR decreases to 1.61. In this case, the effect of residual confounding to increase the estimated OR is initially greater than the attenuating effect of exposure error, causing an increase in the estimated OR. As the correlation between exposure and confounder increases further the relationship is reversed, and the estimated OR decreases.

Table 3.21 and Table 3.22 show the results of simulations in which the ICC of exposure is 0.75 and 0.5 respectively and the correlation between the confounders equals 0.5. When the confounders are correlated, increasing measurement error can decrease the bias in the estimated exposure-outcome OR. The correlation between exposure and confounders can act to decrease the estimated OR. Omitting confounders from the analysis can cause a decrease in the bias of the estimated exposure-outcome OR. All of these effects have been seen previously in Table 3.13 and Table 3.14. The estimated ORs in Table 3.13 and Table 3.14 are generally larger than the true exposure-outcome OR, while the estimated ORs in Table 3.21 and Table 3.22 in general underestimate the true exposure-outcome association.

Table 3.21: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, ICC of exposure = 0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 Crude OR	0.1 OR adjusted for Z_1	0.1 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 … 0.75	0.1 … 1	0.3 Crude OR	0.3 OR adjusted for Z_1	0.3 … 0.5	0.3 … 0.75	0.3 … 1	0.5 Crude OR	0.5 OR adjusted for Z_1	0.5 … 0.5	0.5 … 0.75	0.5 … 1
	0.5		1.80	1.80	1.80	1.80		1.99	1.87	1.80	1.72		2.21	1.95	1.78	1.59
0.1	0.75	1.81	1.80	1.80	1.80	1.80	1.99	2.00	1.89	1.82	1.75	2.19	2.23	1.99	1.83	1.65
	1		1.80	1.80	1.80	1.80		2.01	1.91	1.85	1.78		2.24	2.04	1.89	1.71
	0.5		1.84	1.87	1.89	1.91		2.04	1.95	1.90	1.84		2.28	2.04	1.89	1.73
0.3	0.75	1.99	1.76	1.80	1.82	1.85	2.19	1.97	1.90	1.85	1.81	1.43	2.20	2.00	1.88	1.73
	1		1.68	1.72	1.75	1.78		1.89	1.84	1.81	1.77		2.12	1.96	1.85	1.72
	0.5		1.88	1.95	1.99	2.04		2.10	2.04	2.00	1.96		2.37	2.14	2.01	1.86
0.5	0.75	2.19	1.70	1.78	1.83	1.89	2.43	1.93	1.89	1.88	1.85	2.70	2.18	2.01	1.91	1.79
	1		1.52	1.59	1.65	1.71		1.73	1.73	1.73	1.72		1.98	1.86	1.79	1.69

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=ln2, β_1=β_2=ln2. Width of 95% simulation intervals<0.0073.*

Table 3.22: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, ICC of exposure = 0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.61	1.60	1.60	1.60	1.60	1.73	1.73	1.64	1.59	1.53	1.86	1.87	1.67	1.55	1.42
	0.75		1.60	1.60	1.60	1.60		1.73	1.65	1.60	1.55		1.88	1.70	1.58	1.45
	1		1.60	1.60	1.60	1.60		1.74	1.67	1.62	1.57		1.89	1.73	1.62	1.49
0.3	0.5	1.73	1.62	1.64	1.65	1.67	1.86	1.75	1.69	1.65	1.61	2.00	1.90	1.72	1.62	1.50
	0.75		1.56	1.59	1.60	1.62		1.70	1.65	1.62	1.58		1.85	1.70	1.61	1.50
	1		1.50	1.53	1.55	1.57		1.64	1.61	1.58	1.56		1.79	1.67	1.59	1.50
0.5	0.5	1.86	1.63	1.67	1.70	1.73	2.00	1.77	1.72	1.70	1.67	2.16	1.92	1.77	1.68	1.58
	0.75		1.50	1.55	1.58	1.62		1.64	1.62	1.61	1.59		1.80	1.68	1.61	1.53
	1		1.37	1.42	1.45	1.49		1.51	1.50	1.50	1.50		1.66	1.58	1.53	1.53

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=ln2, β_1=β_2=ln2. Width of 95% simulation intervals<0.0056.*

3.4.2.2 Four confounders

Having considered the impact of exposure measurement error and residual and unmeasured confounding when there are two confounders, the situation with four confounders is now considered.

Table 3.23 and Table 3.24 show the results obtained when the four confounders are uncorrelated, with the exposure ICC equal to 0.75 and 0.5 respectively. The estimated exposure-outcome ORs shown in Table 3.23 are smaller than those obtained when the exposure is measured without error (see Section 3.3.2.2), and the results shown in Table 3.24 are smaller than those in Table 3.23 due to an increase in the exposure measurement error. As the estimated exposure-outcome ORs displayed in Table 3.23 and Table 3.24 are all subject to unmeasured confounding, the estimated ORs are overestimates of the true exposure-outcome OR of two. This contrasts with the results shown in Section 3.4.2.1, where in general the estimated exposure-outcome ORs were underestimates of the true effect.

In Table 3.23 and Table 3.24, increased measurement error does not necessarily lead to an increase in the estimated exposure-outcome association. Consider, for example, Table 3.23, and the scenario in which the correlation between E and X_1 equals 0.1 and the correlation between E and X_2 equals 0.5. When controlling only for Z_1 in the analysis, and the ICC of Z_1 equals one, the estimated exposure-outcome OR is 3.20. When measurement error in Z_1 increases, so that the ICC equals 0.5, the estimated OR decreases to 3.19. This occurs because there is a true exposure-outcome association, as explained in Section 3.3.2.2.

Increased unmeasured confounding may also lead to a decrease in the bias in the estimated exposure-outcome OR. Consider again Table 3.23, and the scenario in which the correlation between E and X_1 is 0.1 and the correlation between E and X_2 equals 0.5. When only Z_1 is controlled for in the analysis, and is measured without error, the estimated OR is 3.20. Increasing unmeasured confounding by omitting Z_1 from the analysis leads to a crude estimated OR of 3.17. Again, this occurs because of the true exposure-outcome association, as described in Section 3.3.2.2.

In Table 3.23, the estimated OR increases as the correlation between exposure and each confounder increases. This is in contrast to the results observed in Table 3.19 where there were only two confounders. As Table 3.23 describes simulations in which there are four confounders, there is a greater residual and unmeasured confounding effect acting to increase the estimated exposure-outcome OR. This is not counteracted by exposure measurement error acting to decrease the estimated OR, and hence a decrease in the estimated OR with an increase in the correlation between the exposure and each confounder is not observed.

Table 3.23: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, ICC of exposure=0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	2.56	2.56	2.55	2.55	2.55	2.84	2.85	2.72	2.66	2.60	3.17	3.19	2.97	2.84	2.71
	0.75		2.56	2.55	2.55	2.55		2.85	2.73	2.66	2.60		3.19	2.97	2.84	2.70
	1		2.55	2.55	2.55	2.55		2.85	2.73	2.66	2.59		3.20	2.97	2.84	2.70
0.3	0.5	2.84	2.73	2.72	2.73	2.73	3.18	3.05	2.92	2.86	2.79	3.58	3.45	3.22	3.09	2.95
	0.75		2.66	2.66	2.66	2.66		2.99	2.86	2.79	2.72		3.38	3.14	3.01	2.86
	1		2.60	2.60	2.60	2.59		2.92	2.79	2.72	2.64		3.31	3.06	2.92	2.77
0.5	0.5	3.17	2.97	2.97	2.97	2.97	3.58	3.37	3.22	3.14	3.06	4.07	3.86	3.61	3.47	3.31
	0.75		2.85	2.84	2.84	2.84		3.25	3.09	3.01	2.92		3.74	3.47	3.31	3.14
	1		2.72	2.71	2.70	2.70		3.12	2.95	2.86	2.77		3.61	3.31	3.14	2.95

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=ln2, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.011.*

Table 3.24: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	2.08	2.08	2.07	2.07	2.07	2.24	2.24	2.15	2.11	2.06	2.41	2.41	2.24	2.15	2.05
	0.75		2.08	2.07	2.07	2.07		2.24	2.15	2.11	2.06		2.42	2.24	2.15	2.05
	1		2.07	2.07	2.07	2.07		2.24	2.15	2.11	2.06		2.42	2.24	2.14	2.04
0.3	0.5	2.24	2.15	2.15	2.15	2.15	2.41	2.33	2.24	2.19	2.15	2.60	2.52	2.34	2.24	2.14
	0.75		2.11	2.11	2.11	2.11		2.28	2.19	2.15	2.10		2.47	2.29	2.19	2.09
	1		2.07	2.06	2.06	2.06		2.24	2.15	2.10	2.05		2.43	2.24	2.14	2.03
0.5	0.5	2.41	2.24	2.24	2.24	2.24	2.60	2.43	2.34	2.29	2.24	2.82	2.64	2.45	2.35	2.24
	0.75		2.15	2.15	2.15	2.14		2.34	2.24	2.19	2.14		2.55	2.35	2.24	2.13
	1		2.06	2.05	2.05	2.04		2.25	2.14	2.09	2.03		2.45	2.24	2.13	2.01

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=ln2, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0068.*

In contrast, Table 3.24 does show a decrease in the estimated OR with an increase in the correlation between exposure and each confounder. Consider, for example, when the correlation between E and X_1 equals 0.1 and the correlation between E and X_2 equals 0.3. When both Z_1 and Z_2 are measured without error the estimated exposure-outcome OR is 2.06. Increasing the correlation between E and X_1 to 0.3 leads to a decrease in the estimated OR to 2.05. A further increase in the correlation between E and X_1 to 0.5 leads to a further decrease in the estimated OR to 2.03. This contrast between Table 3.23 and Table 3.24 is explained by the increase in exposure measurement error between the two tables. While in Table 3.23 the amount of exposure measurement error was not sufficient to counteract the effect of any residual or unmeasured confounding, the greater exposure measurement error in Table 3.24 is able to counteract the effect of any residual and unmeasured confounding in some cases.

Table 3.25 and Table 3.26 show the results when the correlation between the confounders is 0.5 and the exposure ICC is 0.75 and 0.5 respectively. Again, increasing exposure measurement error leads to a decrease in the estimated exposure-outcome OR, which is shown by the fact that the ORs displayed in Table 3.25 are smaller than those observed when exposure is measured without error (see Section 3.3.2.2), and the ORs displayed in Table 3.26 are smaller than those in Table 3.25. In general, the estimated ORs shown in Table 3.25 overestimate the true exposure-outcome OR, while those in Table 3.26 generally underestimate the true exposure-outcome OR.

Table 3.25 and Table 3.26 show that the estimated exposure-outcome OR can decrease with increasing measurement error in the confounders. It is also possible for the estimated OR to decrease with increased unmeasured confounding, and to decrease with increased correlation between exposure and each confounder. These results are unsurprising, as the results of Section 3.3 showed that these trends are the result of correlated confounders.

Table 3.25: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, ICC of exposure=0.75, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 Crude OR	0.1 OR adjusted for Z_1	0.1 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 … = 0.75	0.1 … = 1	0.3 Crude OR	0.3 OR adjusted for Z_1	0.3 … and Z_2, ICC = 0.5	0.3 … = 0.75	0.3 … = 1	0.5 Crude OR	0.5 OR adjusted for Z_1	0.5 … and Z_2, ICC = 0.5	0.5 … = 0.75	0.5 … = 1
0.1	0.5	2.08	2.15	2.22	2.26	2.32	2.25	2.35	2.23	2.16	2.08	2.44	2.58	2.26	2.06	1.83
	0.75		2.20	2.26	2.30	2.36		2.42	2.31	2.23	2.15		2.67	2.38	2.18	1.94
	1		2.27	2.32	2.36	2.41		2.52	2.41	2.34	2.25		2.80	2.53	2.34	2.10
0.3	0.5	2.25	2.12	2.23	2.31	2.41	2.44	2.32	2.24	2.19	2.14	2.66	2.55	2.27	2.09	1.89
	0.75		2.05	2.16	2.23	2.34		2.25	2.19	2.16	2.12		2.49	2.25	2.09	1.91
	1		1.97	2.08	2.15	2.25		2.18	2.14	2.12	2.08		2.43	2.23	2.09	1.92
0.5	0.5	2.44	2.10	2.26	2.38	2.53	2.66	2.31	2.27	2.25	2.23	2.92	2.55	2.30	2.14	1.96
	0.75		1.90	2.06	2.18	2.34		2.10	2.09	2.09	2.09		2.34	2.14	2.02	1.87
	1		1.67	1.83	1.94	2.10		1.88	1.89	1.91	1.92		2.11	1.96	1.87	1.75

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=ln2, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0066.*

Table 3.26: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.79	1.83	1.87	1.90	1.93	1.89	1.95	1.87	1.82	1.76	2.01	2.08	1.86	1.72	1.56
	0.75		1.86	1.90	1.92	1.96		1.99	1.91	1.86	1.81		2.14	1.93	1.79	1.63
	1		1.90	1.93	1.96	1.98		2.05	1.97	1.92	1.87		2.20	2.01	1.88	1.72
0.3	0.5	1.89	1.80	1.87	1.91	1.97	2.01	1.92	1.86	1.83	1.80	2.13	2.05	1.86	1.74	1.60
	0.75		1.75	1.82	1.86	1.92		1.87	1.83	1.81	1.78		2.01	1.84	1.74	1.61
	1		1.69	1.76	1.81	1.87		1.83	1.80	1.78	1.76		1.97	1.83	1.74	1.62
0.5	0.5	2.01	1.76	1.86	1.93	2.01	2.13	1.89	1.86	1.84	1.83	2.27	2.02	1.86	1.75	1.64
	0.75		1.62	1.72	1.79	1.88		1.75	1.74	1.74	1.74		1.88	1.75	1.67	1.57
	1		1.47	1.56	1.63	1.72		1.60	1.60	1.61	1.62		1.73	1.64	1.57	1.50

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=ln2, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0046.*

3.5. Discussion

3.5.1. Summary of results

The validity of an epidemiological study may be threatened by both residual and unmeasured confounding. With plausible assumptions about residual and unmeasured confounding, effect sizes of the magnitude frequently reported in observational epidemiological studies can be generated. This study has shown that if the confounders are uncorrelated, bias in the estimated exposure-outcome ORs increases as error in the measured confounders increases, as the number of unmeasured confounders increases, and as the correlation of the confounders with exposure increases. If the confounders are correlated, or there is a true association between the exposure and outcome, bias in the estimated ORs can decrease as measurement error increases, and as unmeasured confounding increases. Correlated confounders can also cause bias in the estimated ORs to decrease as the correlations between exposure and confounders increases. Unmeasured confounding is a more serious problem when the confounders are uncorrelated, and can result in substantial bias in the estimated exposure-outcome OR, even when only one confounder is omitted from the analysis.

When the effects of exposure measurement error were also considered, in general the trends observed were similar to those observed in Section 3.3 and therefore due to residual or unmeasured confounding. The combined effect of exposure measurement error and residual and unmeasured confounding was more unpredictable when the exposure had a true effect on the outcome, and was able to cause a decrease in the estimated exposure-outcome OR with increased correlation between exposure and each confounder, even when such a decrease was not predicted by residual or unmeasured confounding. One constant trend was observed throughout the investigation of exposure measurement error; the effect of random, non-differential and additive measurement error in the exposure is to attenuate the estimated ORs towards the null value.

3.5.2. Strengths and weaknesses

Only the case in which the errors in the exposure and the confounders are uncorrelated has been considered. Returning to the motivating example of the effect of antioxidant vitamins on cardiovascular outcomes presented in Chapter 1, this assumption may not be realistic. If quantities of nutrients are derived from a questionnaire containing questions about the frequency and quantity of consumption of certain food types, errors in variables may well be correlated. Errors in reporting a food type that contains two nutrients will result in the errors in the quantities of those two nutrients to be correlated.

Here, simulation studies have been used to illustrate the results. While analytic results are

always desirable, they are not always possible to obtain. As stated by Gustafson,[55] *"unfortunately, closed-form expressions for the bias induced by measurement error in logistic regression do not exist"*. The aim here is to examine the effects of unmeasured and residual confounding given parameter combinations representing situations commonly seen in epidemiological research.

Care should be taken in generalizing from the results of a simulation study.[56] Here, it has been assumed that both exposure and confounders were normally distributed with mean zero and variance one. Continuous variables in epidemiological studies may not have a normal distribution, but logarithm or square-root transformations can improve normality (although this can cause added difficulty in interpreting the results of a regression analysis). The results do not depend on the choice of mean and can be interpreted, without loss of generality, as the effect of a single standard deviation increase in the exposure and confounders. Often, in epidemiological studies, some or all of the confounders and/or the exposure are categorical. The results presented here will not generally apply in this situation. Misclassification of categorical (or binary) variables is a more complex problem, as errors will be correlated with the true values.[57] Extensions of the work presented here to deal with categorical exposures and confounders would be desirable.

3.5.3. Implications

Usually, when analyzing an epidemiological study, the true model is not known. We do not reliably know which variables are confounders of the association of interest, the form in which they should enter the model, or the time scale over which they act. It has been suggested that confounders can be identified by evaluating the change in the exposure-outcome estimate.[58] For example, if the estimate adjusted for a variable is more than 15% different from the estimate obtained without adjusting for the variable, then the variable is considered to be a confounder. Strict adherence to such a rule could lead to true confounders being disregarded. Consider, for example, Table 3.1. If the correlation between both E and X_1 and E and X_2 was 0.1, the OR adjusted for each of them separately would only differ from the crude OR by between 2.7 and 5.6 percent, depending on measurement error. This would lead us to believe that X_1 and X_2 were not confounders, and that the crude OR was the true exposure effect estimate. Rules such as the change in estimate criterion should be applied carefully, or not at all.[59]

This chapter highlights the need to perform sensitivity analyses to assess whether unmeasured confounding is a likely problem. Unmeasured confounders have been shown to have a cumulative effect on the bias of exposure effect estimates. The possibility of several unmeasured confounders should be taken into account when performing sensitivity analyses. It may not be enough to state that a single unmeasured confounder would need an implausibly large OR to remove the observed effect. Several unmeasured confounders with small or

moderate effects may be able to produce the same effects. Sensitivity analysis methods to assess the possible effects of selection bias, misclassification of covariates and unmeasured confounding have been proposed and illustrated by, for example, Greenland[60] and Lash and Fink.[61] Sensitivity analysis methods will be discussed in Chapter 4.

If information is available on confounders, it should be used in the estimation of effect estimates. For example, Khaw, Bingham, Welch *et al.*[5] note in their discussion that while information on social class and physical activity of the participants was recorded, they were not used in the reported analysis. If these variables were indeed confounders of the relationship between ascorbic acid and mortality, even a moderate effect of each would result in sizable residual confounding in the reported estimates, as shown by Lawlor, Davey Smith, Bruckdorfer *et al.*[11] Confounders may be omitted from the analysis because of missing data leading to loss of information. In these situations, missing data methods (e.g. multiple imputation[62, 63]) can be used to avoid bias due to unmeasured confounding.

When there is a true association between exposure and outcome, the true effect can be either underestimated or overestimated, depending on the amount of residual and unmeasured confounding. Investigators should therefore take care about stating that any exposure measurement error will have resulted in a conservative estimate of the exposure effect without considering the additional effect of measurement error in confounders, or of confounders omitted from the analysis. As shown in this chapter, the effects of measurement error in explanatory variables may not be easy to predict. If no residual or unmeasured confounding exists, exposure error of the type considered here will cause an underestimate of the true association, in which case investigators would be correct to state that the observed estimate is conservative.

The effect of measurement error on exposure effect estimates should be explored, either by adjusting the estimates based on knowledge of the likely measurement error, or by performing sensitivity analyses. Of course, the ideal is that the variables are measured without error, but this is unlikely to occur in reality. While effort should be used to minimize the measurement error that occurs, evaluation of the measurement error that has occurred should be quantified and used in the final estimate. A wide variety of methods are available to correct exposure effect estimates for measurement error (see Chapter 4).

3.5.4. Future research

As an extension to the results presented in this chapter, further simulation studies could be carried out to investigate the effects of systematic and/or differential measurement errors on estimated exposure-outcome associations. An investigation of the effect of measurement errors in non-normally distributed continuous explanatory variables, or misclassification of

explanatory variables could be carried out. Additionally, the effects of measurement errors that act multiplicatively on the true variable of interest, or errors that are correlated with other errors or the true underlying variable could be investigated.

A further issue when considering the effects of measurement errors in explanatory variables in epidemiological studies is recovering the true exposure effect estimate, given information about the amount of measurement error or misclassification in the explanatory variables. In Part B, the results presented in this chapter are extended by correcting the exposure effect estimates for measurement error.

Part B.

Allowing and correcting for measurement error in epidemiological studies

BLANK IN ORIGINAL

Chapter 4.

Background literature: Measurement error correction and sensitivity analysis methods in epidemiology

4.1. Introduction

As shown in Chapter 3, measurement error in epidemiological studies can bias the estimates of the exposure-outcome association, and lead to incorrect conclusions about the size or direction of the association. Measurement error adjustment methods have been devised to combat this problem. These methods generally assume that an internal or external validation study, or replicate measures of the variables measured with error exist to enable estimation of the measurement error parameters. Section 4.2 briefly describes methods to estimate measurement error parameters. Some methods to correct for measurement error in epidemiological studies are described in Section 4.3. If the measurement error or misclassification parameters are not known and cannot be estimated, sensitivity analysis can be used to investigate the possible impact of various values of the mismeasurement parameters on the effect estimates from epidemiological studies. Sensitivity analysis methods are described in Section 4.4.

4.2. Estimation of measurement error or misclassification parameters

In many epidemiological studies, regression analysis is used to estimate exposure effects. Consider the simple example in which outcome is binary, and there is a single continuous exposure variable. Logistic regression is used to estimate the exposure-outcome OR, and the logistic model is

$$\ln\left(\frac{\pi}{1-\pi}\right)=\alpha+\beta_E E$$

Here, π is the probability of the outcome, E is the exposure variable measured without error, and β_E is the parameter of interest, the exposure-outcome log OR. If the exposure is measured with error as Z_E, and a classical measurement error model is assumed, then

$$Z_E=E+\varepsilon$$

where ε is the measurement error variable, and is assumed to have mean zero and variance σ_e^2. The logistic regression model using the available data is therefore

$$\ln\left(\frac{\pi}{1-\pi}\right)=\alpha+\beta_E^* Z_E$$

In general, $\beta_E^* \neq \beta_E$, and in order to obtain an estimate of β_E the measurement error parameters must be estimated. In the case of continuous variables measured with error, the measurement error parameter is usually the variance of the errors, which can be estimated using, for example, the intra-class correlation coefficient (see Section 3.2.2). For misclassified categorical variables, the measurement error parameters of interest are usually the probabilities of misclassification. In the special case of binary variables, there are two misclassification probabilities of interest, known as the sensitivity and specificity.

4.2.1. Known measurement error parameters

The simplest case occurs when the measurement error parameters are known. In this situation, there is no need for any parameter estimation. For example, in the simulation studies presented so far, the measurement error variance is a parameter defined in the simulation stages, and thus is known. This means that, to correct the residual confounding in the simulations presented in Chapter 3, the known measurement error variance can be used. This situation is unlikely to occur in epidemiological studies, and the measurement error parameters will usually need to be estimated.

4.2.2. Validation study

In validation studies, the variable is measured without error using a gold-standard measurement process, and the variable measured with error is also recorded. There are two types of validation studies. In internal validation studies, the perfectly measured variable is recorded on a subset of the participants in the main study. External validation studies involve a sample of the population not included in the main study. The measurement error parameters, such as the error variance, or sensitivity and specificity, relating the true and error-prone measures can then be estimated.

4.2.3. Replicate measures

Certain variables in an epidemiological study may be measured more than once, at different points in time. These replicate measurements can provide an estimate of the measurement error variance, using a one-way analysis of variance, the Pearson correlation coefficient with each pair of measurements entered twice, once in reverse order, or a simple random effects model.[44] Misclassification probabilities can also be estimated from replicate measures.

4.2.4. Transportation

If validation study data or replicate measures are not available, parameters from other studies may be used to estimate the measurement error parameters. If the same measurement error model does not hold for the two populations, this method carries a risk of introducing bias into the main regression analysis.

4.3. Measurement error correction methods

While it is important to recognise the potential effects of measurement error in explanatory variables on estimated exposure-outcome associations in epidemiology, it is desirable to be able to correct the estimates. This has received considerable attention in the literature, with many methods available. Much is concerned with studies in which validation data are available, and therefore the measurement error parameters can be estimated. In this section, some methods for measurement error correction are reviewed. As there is a large amount of literature on this subject, the review here does not include all available publications. Further references can be

found from other published reviews of measurement error correction methods. For example, Thomas, Stram and Dwyer[64] review methods for correcting for exposure measurement error. Carroll, Ruppert and Stefanski[65] review methods available for measurement error correction in non-linear models. Thürigen, Spiegelman, Blettner *et al.*[66] provide a review of correction methods applicable to case-control studies. Bayesian methods for measurement error correction are reviewed by Gustafson.[55]

The following literature is organised into five categories. The first two categories describe functional modelling and structural modelling methods (Section 4.3.1 and Section 4.3.2 respectively). The definitions of these methods are the same as those used by Carroll, Ruppert and Stefanski.[65] Suppose that **X** is the vector of the exposure and confounders of interest, which is measured without error and unobserved. Functional models make few, if any, assumptions about the distribution of **X**, and therefore do not parametrically model the distribution of **X**. In contrast, the distribution of **X** is parametrically modelled in structural models. Functional modelling is appealing in situations in which measurements of the exposure and confounders are made only on the units of interest, rather than on a random sample of a population. In this situation, the true values of the exposure and confounders can be considered unknown constants, and no assumptions about their distribution are required. Alternatively, if measurements are made on a random sample from a population, it would be desirable to model the distribution of the exposure and confounders in that population, and therefore adopt a structural modelling approach. Functional modelling is still useful where the exposure and confounders are random variables, as the lack of assumptions made about their distributions makes functional modelling approaches robust to misspecification of the distribution models.

The third section considers Bayesian methods (Section 4.3.3). These approaches define a prior distribution for the parameter of interest, based on available knowledge about its distribution. From Bayes theorem, the posterior distribution of the parameter given the observed data is proportional to the product of the model for the observed data given the parameter and the prior distribution. The point estimate can be defined as, for example, the mean, median or mode of the posterior distribution, and a (100-α)% credible interval, which is analogous to a frequentist confidence interval, is defined as the interval with probability (1-0.01α) under the posterior distribution. The posterior distribution can be difficult to compute, and numerical integration methods, such as Markov Chain Monte Carlo algorithms, are often used.

The fourth section describes matrix methods for obtaining parameter estimates when categorical variables are misclassified (Section 4.3.4). In principle, these are simple to apply, and are based on the idea that the observed matrix of data is equal to the true matrix of data multiplied by the matrix of misclassification probabilities. The true data is then obtained by multiplying the observed data by the inverse of the matrix of misclassification probabilities, and

used to obtain effect estimates. The final section describes methods for correcting for measurement error which do not fit into any of the other categories (Section 4.3.5).

4.3.1. Functional modelling methods

4.3.1.1 Regression calibration

In regression calibration models, the relationship between the error-prone variables and the error-free variables is used to either estimate the values of the unobserved true variables to be used in the main regression analysis, or to correct the parameter estimates obtained from the main regression analysis in which the error-prone variables are included as covariates.

A regression calibration method was described by Prentice[67] to correct estimated regression parameters in Cox proportional hazards models for measurement error. In this method, the regression of the unobserved perfectly measured covariates on the observed, error-prone covariates and observed perfectly measured covariates is used to correct the parameter estimates from the main regression model for measurement error. Although confidence interval estimation was not considered by Prentice,[67] later work on regression calibration has included methods for estimating confidence intervals that can be used in this situation. Rosner, Willett and Spiegelman[68] and Rosner, Spiegelman and Willett[69] developed a regression calibration estimator in the logistic regression setting that is equivalent to the estimator described by Prentice.[67]

A second regression calibration method was described by Carroll and Stefanski.[70] This method was further described in Carroll, Ruppert and Stefanski,[65] and differs slightly from the approach described by Rosner *et al.*[68, 69] Rather than correcting the parameter estimates from the main regression model, Carroll and Stefanski[70] estimated the values of the true variable in the main dataset, and use this estimated variable in the main regression analysis. This regression calibration method is described in more detail in Section 5.2.4. The two regression calibration methods described by Rosner *et al.*[68, 69] and Carroll and Stefanski[70] have been shown to be equivalent under fairly general assumptions in main study/external validation study designs by Thurston, Spiegelman and Ruppert.[71]

Extensions to the basic regression calibration method of Rosner *et al.*[68, 69] have been proposed. Kipnis, Carroll, Freedman *et al.*[72] described an extension to deal with errors that are correlated with the true values of a variable, and correlated systematic errors. Special cases of this model have been previously proposed by Freedman, Carroll and Wax,[73] and Spiegelman, Schneeweiss and McDermott.[74] A problem with this regression calibration method is that the parameters cannot all be identified by standard validation study designs. This problem was considered by Spiegelman, Zhao and Kim,[75] who described study designs and estimation methods that allow

the model parameters to be estimated. A further extension to the basic regression calibration model of Rosner *et al.*[69] to include studies with internal validation designs has been described by Spiegelman, Carroll and Kipnis.[76]

A regression calibration method for logistic regression and case-control data was described by Armstrong, Whittemore and Howe.[32] This method assumed a normal discriminant analysis model for the measurement error, and also allows for differential measurement errors. The method is similar to that of Rosner *et al.*,[68, 69] in that the naïve estimates are corrected for measurement error using estimates of the covariance matrices of the unobserved, perfectly measured variables and the error variables. These estimates can be obtained from repeated measures. The variance formulae provided for the corrected regression parameters do not account for the additional variability caused by estimation of the covariance matrices.

Tosteson, Stefanski and Schafer[77] described a regression calibration method for probit models. This method assumed a Berkson measurement error model, and the parameters of the measurement error model were assumed to be known. In reality, these parameters will usually be estimated from a validation study. The additional variability of the corrected probit regression parameters caused by estimation of the measurement error model parameters was not considered.

Gleser[78] proposed a regression calibration approach for linear regression analyses with errors in continuous covariates. In this method, the reliability matrix of the covariates was estimated using a validation study or replicate measurements, and used to transform the observed covariates. These transformed covariates were then used in a linear regression model to estimate the parameters of interest. This approach is similar to that of Carroll and Stefanski.[70]

Xie, Wang and Prentice[79] also described a regression calibration estimator for use in a Cox proportional hazards model. Using replicate measures, and assuming a classical measurement error structure, the model was recalibrated at each distinct failure time. This method was extended by Gorfine, Hsu and Prentice[80] to account for stratified Cox proportional hazards regression models. This extension did not perform well in simulations in which the covariates were non-normally distributed.

White, Frost and Tokunaga[57] considered a regression calibration approach to non-differential measurement error in both binary and continuous variables, using replicates to estimate the measurement error probabilities. They assumed that the outcome was continuous, and the mismeasured continuous variables were assumed to be uncorrelated with their errors. For binary variables, the measurement error probabilities were unidentified if only two replicates were available. This is not a concern if the binary variable is a confounder, but if it is the

exposure either three replicates or further assumptions are required for identifiability. Although the outcome variable was assumed to be continuous, the method would approximately apply to other generalized linear models.

4.3.1.2 Simulation extrapolation

Simulation-extrapolation is a method for correcting for measurement errors in continuous variables, and has been described by Cook and Stefanski,[81] Carroll, Kuchenhoff, Lombard, *et al.*,[82] and Carroll, Ruppert and Stefanski.[65] The method involves adding further measurement error to the observed error-prone variables, estimating the exposure-outcome association for several values of additional measurement error, and extrapolating the true exposure-outcome association from the observed relationship between the estimated exposure-outcome association and measurement error. Three extrapolant functions are generally considered. These are linear, quadratic, and rational linear (i.e. a ratio of two linear expressions) functions of the amount of measurement error. Cook and Stefanski[81] showed that using a linear or quadratic extrapolant generally produces conservative corrections for measurement error. The rational linear extrapolant may produce better corrections for measurement error, but it can be numerically unstable when the effects of measurement error on the parameter of interest are small, and therefore the extrapolant is an almost horizontal line. It may also produce singularities in the region of measurement error between the observed parameter and the extrapolated parameter corrected for measurement error.[65] The simulation-extrapolation method is described in more detail in Section 5.2.3. A major advantage of the method is that it is not restricted to a particular regression model. The method is, however, computationally intensive.

Fung and Krewski[83] used a simulation study to compare a regression calibration method with simulation extrapolation when a single variable is measured with error in a logistic regression analysis. They simulated an exposure and a single confounder, and only one of these was measured with error. They found that the regression calibration method worked well when the measurement error was additive, or for Berkson measurement error when the correlation between the exposure and confounder was close to zero. The method did not perform as well when the measurement error was Berkson and the exposure and confounder were highly correlated. Simulation-extrapolation was not as successful as regression calibration at removing bias due to measurement error in the effect estimates.

4.3.1.3 Semi-parametric methods

In semi-parametric methods, the measurement error model is partially or fully unspecified. These methods may involve non-parametric estimation of the distributions of the perfectly measured or error-prone covariates, non-parametric estimation of the moments of the conditional distributions of the perfectly measured or error-prone covariates, or non-parametric estimation of the likelihood.

A semi-parametric approach to correcting for measurement errors in covariates has been described by Robins, Rotnitzky and Zhao[84] and Robins, Hsieh and Newey.[85] This approach involves a parametric model for the relationship between outcome and the explanatory variables, but no parametric assumptions are made about the measurement error model. The proposed estimator is consistent and semi-parametric efficient. This semi-parametric approach can be used in a variety of epidemiological situations, as it can account for both continuous and categorical outcomes, validation studies that are not randomly selected, and any number of mismeasured covariates. In addition, the method does not require the assumption that measurement error is non-differential. Spiegelman and Casella[86] demonstrated that the expressions provided by Robins, Hsieh and Newey[85] simplify if the outcome variable is binary, the validation sample is randomly selected, and there is one covariate measured with error.

Simulation studies were used by Stürmer, Thürigen, Spiegelman *et al.*[87] to compare the performance of Rosner *et al.*'s[68, 69] regression calibration model, and the semi-parametric method for correcting for measurement error described by Robins, Hsieh and Newey.[85] The results of the simulation studies showed that the semi-parametric approach was affected less by large error variances and differential errors than the regression calibration estimate. The semi-parametric approach is, however, complex. In situations in which the assumptions of the regression calibration model hold, investigators may prefer to use this intuitively and computationally simpler method.

A semi-parametric mixture model was proposed by Roeder, Carroll and Lindsay[88] to account for misclassification in case-control studies. Although the method was developed in terms of misclassification of a categorical variable, the results extend to the case of a continuous covariate. The method was developed in settings with internal validation studies, but can also be used if external validation data is available, as long as the measurement error is non-differential with respect to the outcome. The method can also account for additional perfectly measured covariates, but was developed in the situation where only one covariate is measured with error.

Gorfine, Hsu and Prentice[80] proposed a non-parametric estimator to correct for measurement error in Cox proportional hazards model. Using simulation studies, their estimator was shown to substantially reduce bias when the covariates were non-normally distributed. This robustness to the measurement error model may be important in practical applications, where the true measurement error model is not known.

4.3.1.4 Other functional modelling methods

A functional method involving correcting the partial likelihood score function[89] in Cox proportional hazards models was originally proposed by Nakamura.[90] This method assumes

additive and normally distributed measurement errors with a known covariance matrix. In practical applications, the covariance matrix of the errors may not be known, and will have to be estimated. The method was extended by Huang and Wang[91] to the situation in which the measurement error parameters are estimated using replicated data. This extension only assumes additive measurement errors, and does not make any further assumptions about the distributions of the covariates or measurement errors. Another extension to Nakamura's[90] method was proposed by Hu and Lin,[92] where the measurement error parameters are estimated using either a randomly selected validation sample, or replicated measures. In this extension, no assumptions are made about the distributions of the covariates, but the errors are assumed to have a symmetric distribution.

Gorfine, Hsu and Prentice[80] extended two functional methods[91, 92] for use with stratified Cox regression models. The extensions of Huang and Wang's[91] method, and Hu and Lin's[92] method tended either not to converge, or to be highly biased even when the measurement error variance was small.

Measurement error in continuous explanatory variables and matched case-control studies was considered by McShane, Midthune, Dorgan, *et al.*[93] A conditional scores method for logistic regression analysis was developed assuming normally distributed, non-differential measurement errors. Using simulation studies, the conditional scores method was compared with a regression calibration method and was shown to be superior when the explanatory variables were non-normally distributed and highly skewed.

4.3.2. Structural modelling methods

A pseudolikelihood method was proposed by Carroll, Gail and Lubin[94] to deal with measurement errors in covariates in case-control studies. They used a prospective logistic regression model and a parametric measurement error model to express the retrospective likelihood. Pseudolikelihoods were obtained by substituting the true values of the parameters of the measurement error model with estimates of the parameters, and pseudolikelihood estimates of the regression parameters of interest were acquired. The method can account for differential errors, and for errors in both categorical and continuous covariates, and was developed in the setting of an internal validation study.

Measurement error in covariates in case-control studies and retrospective logistic regression was considered by Forbes and Santner.[95] Three estimators of the conditional maximum likelihood were presented that used different information about the measurement error model. The first estimator was a bias-corrected version of the uncorrected conditional maximum likelihood estimator (MLE) based on its asymptotic bias. For the second estimator, the measurement errors were assumed to be normally distributed. The third estimator was based

on score equations in which the measurement error was not assumed to be normally distributed. The properties of the three estimators were investigated using simulation studies, and the bias-corrected estimator was found to perform best in the presence of measurement error.

In segmented regression, the analytical form and parameters for the model relating the exposure and confounders to the outcome may be different for different values of the explanatory variables. Threshold models are a special case of segmented regression. Küchenhoff and Carroll[96] considered measurement error in explanatory variables in segmented regression analyses. They described a MLE that assumed a parametric distributional form for the unobserved, perfectly measured covariate. They found that the MLE, when the model was correctly specified, was much less variable than regression calibration or simulation-extrapolation estimates. When the model was misspecified, the MLE did not perform well.

Measurement error in a single continuous covariate in a Cox proportional hazards model was considered by Hu, Tsiatis and Davidian.[97] Their method used a likelihood-based approach. This approach has been extended by Liu, Mazumdar, Stone *et al.*[98] to include an additional, perfectly measured binary variable. The extension also allows measurement errors to be differential with respect to the binary covariate, and allows for correlations between repeated measures of the error-prone variable.

Kosinski and Flanders[99] proposed a method for correcting for misclassification of a binary exposure in which a gold standard measurement is not required, and instead two imperfect test results are available. The method models the probabilities of exposure, test one, and test two by logistic regression analyses, and can accommodate both differential and non-differential misclassification errors. An expectation-maximisation (EM) algorithm is used to obtain the maximum-likelihood estimates of the required parameters. The odds ratio is then approximated using the ratio of exposure odds for diseased subjects relative to undiseased subjects. The method allows for additional covariates, and is easily implemented in standard statistical packages.

Spiegelman, Rosner and Logan[100] proposed a method for estimation of logistic regression parameters that corrects for both measurement error in continuous covariates and misclassification of categorical covariates. Measurement error in the continuous covariates was assumed to have a multivariate normal distribution, and the misclassification of the categorical variables was described using a chain of logistic regression models. Maximum likelihood estimates were obtained for the parameters of the logistic regression model in the main study and the measurement error and misclassification models. Simulation studies were used to compare the MLE with a regression calibration estimate. These studies showed that regression

calibration may be preferable for situations in which the validation study size is between 100 and 200 and the true regression parameter of interest is close to the null. The MLE performed better than regression calibration when the value of the true parameter was not close to the null, and data from an internal validation study was available. The MLE did not perform well when validation data came from an external study, and in these situations regression calibration may be a preferable approach.

A method for correcting for misclassification of a binary outcome and measurement error in a single covariate has been described by Roy, Banerjee and Maiti.[101] The method was developed for both probit and logistic regression models, and the parameters of the models were estimated using maximum likelihood methods. Simulation studies showed that the model works fairly well for correcting for misclassification of outcome and measurement error in a single covariate. Measurement errors in multiple covariates cannot be corrected for using the described method, and the method assumes that measurement errors and misclassifications are non-differential.

4.3.3. Bayesian methods

Bayesian methods have been proposed to account for measurement error in epidemiological studies. Several of these methods apply to analysis of case-control studies.

Müller and Roeder[102] described a Bayesian semi-parametric approach for case-control studies with errors in variables. Measurement error was assumed to be non-differential and to only affect a single explanatory variable. A mixture of multivariate normal models for the covariates was assumed, with a Dirichlet process prior model on the unknown mixture measure. This method is computationally complex. Using multivariate normal kernels implies continuous covariates, and extensions to the model would be required for categorical covariates.

A method for correcting for measurement error in the exposure in unmatched case-control studies was described by Gustafson, Le and Vallée.[103] They used a normal discriminant model for the unobserved true exposures, and assumed that measurement error was non-differential. The method allows for any number of explanatory variables. The assumption of a normal discriminant model means that the exposure variables must be continuous, and that the model cannot be used in situations with categorical explanatory variables.

Another Bayesian method for use in case-control studies with errors in covariates was described by Gustafson, Le and Vallée.[104] The conditional variance of the error-prone variable, Z, given the perfectly measured variable X and the outcome Y was assumed not to vary with Y. This is a weaker assumption than that of non-differential errors, which assumes that the conditional distribution of Z given X and Y does not depend on Y. The method involves assuming that the distribution of Z is discrete, using a support grid with grid points chosen according to the study

design. Markov Chain Monte Carlo methods are then used to sample from the distribution, and an importance weighting scheme used for parameter inference. The method was formulated using a single mismeasured exposure, and more work is required on the best way to include additional precisely measured covariates into the analysis.

Exposure misclassification in matched case-control studies with a variety of matching ratios was explored by Rice.[105] The method can be viewed as either a random effects model, or as a Bayesian method. In the simplest case of 1:1 matching, the method uses error matrices to correct the observed data, and then the MLE to find the point estimate. For more complex data structures, the Bayesian method is simpler to implement. Only a single exposure variable was considered, which therefore limits the method's applicability to more general epidemiological studies where confounding is likely.

Prescott and Garthwaite[106] proposed three methods for misclassified binary data from a matched case-control study with a validation sub-study. The first two models proposed examine the data in two stages. The first stage combines the data from the validation sub-study with a non-informative prior distribution. The resulting posterior distribution is used as the prior distribution for the second stage. This is then combined with the main study data. The third model uses a hierarchical structure to model the relationship of the exposure probabilities of the matched sets. Much of the paper was concerned with the situation in which each case is matched with a single control, although an extension to multiple matched controls for each case was considered. Throughout, only misclassification in the binary exposure variable was considered, and the models only included the outcome and exposure, with no confounders. This situation is highly unlikely in observational epidemiology, although the authors state in the discussion that the models could be extended to accommodate binary or categorical perfectly measured confounders.

There have also been applications of Bayesian methods to measurement error problems outside of the case-control setting.

Richardson, Leblond, Jaussent *et al.*[107] considered studies in which there is a validation sub-study. They assumed that a single explanatory variable was measured with error, and that the error was non-differential. Their emphasis was on specifying the prior distribution for the analysis, for which they used a mixture of normal distributions with an unknown number of components. Markov Chain Monte Carlo methods were then used for estimation of the model parameters.

In Chapters 4 and 5 of his book, Gustafson[55] described Bayesian methods for correcting for measurement error in continuous and categorical variables, respectively. Methods for both

prospective (e.g. cohort studies) and retrospective (e.g. case-control studies) analyses were considered. The emphasis was on a single mismeasured explanatory variable, and the chapter on misclassification of categorical variables only considered the special case where the mismeasured variable was binary.

4.3.4. Matrix methods

Methods for adjusting for measurement error in continuous variables, and those for adjusting for misclassification in categorical variables are often quite different. One reason for the difference is that, in categorical variables, the error is correlated with the true value of the variable. As explained by White, Frost and Tokunaga,[57] with binary variables if the true value of the variable is zero, then the error is zero or one. If the true value of the variable is one, the error is either zero or -1. The error is therefore negatively correlated with the true value. Matrix methods, which use the misclassification matrices, are simple methods to correct for misclassification in categorical variables.

Much of the earlier work on misclassification of discrete data focussed on 2x2 contingency tables. Barron[108] described a simple method for correcting relative risks from 2x2 tables, using the observed classification matrix, and the two matrices of the conditional probabilities of misclassifying each element in the observed 2x2 table. This method assumed that misclassification was non-differential.

Brenner[27] described a similar matrix method to correct for non-differential exposure misclassification when trend is being assessed. In the method, exposure was categorical and could have more than two levels. The observed matrix of numbers of cases and controls for each exposure level was corrected for misclassification by post-multiplying by the inverse of the misclassification matrix.

Copeland, Checkoway, McMichael *et al.*[109] described a method to correct observed relative risks for exposure misclassification using the sensitivity and specificity of exposure classification. The method can be applied to data from cohort studies and unmatched case-control studies. The exposure was assumed to be binary, and it was assumed that there are no confounders of the exposure-outcome association. It was additionally assumed that the misclassification was random, i.e. that the sensitivity and specificity of exposure classification were the same for both the exposed and unexposed populations. Variance estimation, and therefore correcting 95% CIs for exposure misclassification, was not considered.

An extension to Copeland, Checkoway, McMichael *et al.*'s[109] method to matched case-control studies was described in Greenland[110] and Greenland and Kleinbaum,[111] using the sensitivity, specificity, false positive rate and false negative rate among cases and control separately.

Greenland and Kleinbaum[111] also considered differential misclassifications as an extension to Barron's[108] method. Variance estimation for the corrected estimates was considered by Greenland.[112] Duffy, Rohan, Kandel *et al.*[113] extended this methodology[110-112] to the situation of variable and multiple controls per case. All of these methods considered a single binary exposure, although extensions to accommodate more than one explanatory variable, and more general categorical variables are possible.

Espeland and Hui[114] proposed an approach for analysing 2x2 tables that are subject to misclassification. The method is more complex than previously proposed methods for misclassification of 2x2 tables, but provides variance estimates that are often not available in the simpler methods. Log-linear models are used to describe the misclassification, and a gold standard measurement of the variable of interest is required. The method can be extended to polytomous variables, and to situations with more than two variables. If the misclassification cannot be described by a log-linear model, as would be the case, for example, with misclassification in polytomous variables in a matched case-control study, the proposed method cannot be used.

Weinkam, Rosenbaum and Sterling[115] considered obtaining corrected relative risk estimates when a categorical exposure and a single confounder may be misclassified. The misclassification was assumed to be non-differential. They used the misclassification matrix to recover the true relative risk estimates. Unlike many other methods for misclassified data, they demonstrated their method on categorical data with more than two levels.

One general problem in using matrix methods to correct for misclassification is that the misclassification matrix may be singular, and therefore not invertible. This problem is unlikely to occur in reality unless the classification process is very unreliable.[111]

4.3.5. Other methods

A method for correcting exposure-outcome ORs for non-differential misclassification in a binary confounder was described by Savitz and Barón.[116] They focussed on the case with binary exposure and a single binary confounder. The correction was based on the crude observed OR, the OR adjusted for the misclassified confounder, and estimates of the sensitivity and specificity of the confounder misclassification. The method does not account for sources of bias other than from the misclassification of the single confounder, which may limit the applicability of the method in practice. In addition, the method used simulated data to estimate the effect of sensitivity and specificity on the bias in the observed OR. If the assumptions used for the simulations do not hold for a real-life application, the correction method may not perform well.

Liu and Liang[117] considered non-differential misclassification in generalized linear models. They focussed on the case where multiple observations of a surrogate of the variable of interest were available. These can be viewed as replicate measures of the true variable subject to misclassification. The method used latent class analysis, with an EM algorithm to estimate the regression parameters. The method can only be applied to generalized linear models in which all explanatory variables are categorical. A further possible problem with the method is that the number of parameters that must be estimated increases as the number of variables measured with error, and the number of categories in those variables, increases. This method has been further extended to the situation in which data is collected through two-stage sampling by Emsley, Gao, Hall *et al.*[118]

A method for correcting for misclassification of a categorical exposure was described by Reade-Christopher and Kupper.[119] The method modelled disease risks using logistic models, and disease rates using log-linear models. The method accounted for an exposure with any number of categories, and any number of additional categorical confounders. Misclassification of confounders was not accounted for.

Wang and Pepe[120] described a method to allow for measurement error in covariates using expected estimating equations (EEE). The method can be used when replicate measurements on the variable measured with error are available. The EEE estimator is equivalent to the MLE when the score equations are derived from the likelihood, and conditional expectations are conditioned on the complete dataset. A pseudo-EEE estimator was also developed to reduce the computational complexity of the EEE estimator. Simulation studies were used to evaluate the performance of the EEE and pseudo-EEE estimators, and compared with the performance of a regression calibration estimate. These simulations showed that if the relationship between the covariate measured with error and outcome was not too large, and the sample size was moderate, the regression calibration estimate performed the best. For larger sample sizes and covariate effects, the EEE estimator performed the best.

4.3.6. Summary

This review has shown that there are many methods available to correct observed exposure-outcome effects for the effects of measurement error in exposures and confounders. These methods are not routinely used in the analysis of epidemiological data, where it is likely that some, or even all, of the explanatory variables will be measured with error. Choosing which of these methods to use in a practical application will generally depend on the data in question.

The assumptions for each of the models should apply to the data to be analysed. The regression calibration methods of Rosner *et al.*[68, 69] and Carroll and Stefanski[70] require the exposure and confounders to have a multivariate normal distribution. If this is not the case, other methods

may be more appropriate. Many of the methods assume non-differential measurement error, which may not be an appropriate assumption in, for example, case-control studies where there may be recall bias. In this situation, methods that have been developed to allow for differential measurement error, such as the method described by Armstrong *et al.*,[32] should be used. Structural modelling methods make assumptions about the distribution of the perfectly measured variables, which should be checked. The use of functional modelling methods reduces the number of assumptions made. Violations of the model assumptions may lead to bias in the corrected estimates, and therefore should be considered carefully.

A further consideration is the applicability of the method to the data. Some of the methods are suitable for only continuous, or only categorical data. In situations in which both types of data are present, methods which accommodate both are appropriate. In addition, several methods allow only for measurement error in a single variable. If there are multiple variables measured with error, as is likely in an epidemiological study, these methods will not be appropriate. It may also be important for the method to allow for additional, perfectly measured variables. Age, for example, is commonly included as a confounder in analyses of epidemiological studies, and is generally considered to be measured without error. The form of the outcome variable, and therefore the disease model, should also be considered. Some of the methods proposed are restricted to a particular type of outcome variable, such as a survival time outcome, or a continuous outcome. The measurement error correction method chosen should apply to the outcome in question. Several methods, such as regression calibration, simulation-extrapolation, and the semi-parametric approach of Robins *et al.*,[84, 85] allow for several types of outcome.

Ease of implementation may also be a consideration when choosing a measurement error correction method. For example, programs to implement regression calibration methods are available for Splus, SAS and Stata. Difficulties with programming for the more complex methods may be the main barrier to their implementation. Thürigen, Spiegelman, Blettner *et al.*[66] suggested that there are two conflicting aspects to consider when choosing a measurement error correction method. Methods that are easy to use may not provide estimates that are as precise as other methods, but methods that are theoretically precise may not be easy to implement.

The motivation for the literature review presented in this section is to find methods to correct the estimates obtained from some of the simulation studies in Chapter 3 for measurement error. As the exposure and confounders in the simulations were all continuous, the matrix methods described in Section 4.3.4 are not applicable. The simulation studies investigate the impact of measurement error in the exposure and confounders in a frequentist logistic regression model, and therefore Bayesian methods to correct for measurement error will not be considered. Additionally, using a Bayesian method to correct for measurement error in the simulation

studies would be extremely computationally intensive. The main concern with structural modelling methods is misspecification of the distribution of the unobserved, perfectly measured variables. Although this will not be a problem if using a structural modelling method to correct for measurement error in the simulation studies, as the distributions of all variables are known, there is a possibility of model misspecification when using structural modelling methods for real epidemiological data. Although there is still a possibility for model misspecification for functional modelling methods, fewer assumptions are made and therefore there is less scope for misspecification. Functional modelling methods, and in particular regression calibration and simulation-extrapolation, will therefore be used in Chapter 5 to correct for measurement error in the exposure and confounders in the simulation studies. These two methods are intuitively simpler than many of the other available methods, are applicable to a range of regression models, and are available in several commonly used statistical programs (e.g. SAS and Stata). Carroll, Ruppert and Stefanski[65] have termed these methods the *"default approaches"* (page xviii) to measurement error correction problems.

4.4. Sensitivity analysis methods

If the measurement error variance or misclassification parameters cannot be estimated using validation studies or replicate measures, sensitivity analysis methods can be used to investigate the possible impact of measurement error and misclassification on exposure-outcome effect estimates. Sensitivity analysis can also be used to investigate the possible impact of unmeasured confounding and selection bias on effect estimates.

4.4.1. Methods for unmeasured confounding

One of the first examples of using sensitivity analysis in epidemiology was given by Cornfield, Haenszel, Hammond, *et al.*[121] They used sensitivity analysis to show that the observed association between cigarette smoking and lung cancer was unlikely to be due to unmeasured confounding.

Bross[122] developed a method to determine whether an observed exposure effect is due to a single binary unmeasured confounder. Let Y denote the binary outcome, E denote the binary exposure, and X denote the binary unmeasured confounder. The size of the association between the X and Y is estimated as

$$U^* = \frac{(A+V)(B+1)}{(B+V)(A+1)}.$$

In this formula, A is the ratio of the expected number of people with E=1 and X=0 to the expected number of people with E=1 and X=1. B is the ratio of the expected number of people with E=0 and X=0 to the expected number of people with E=0 and X=1, and V is the confounder-outcome risk ratio. Estimates of A, B and V will need to be obtained from external data. If the size of U^* is not as large as the observed effect of exposure on outcome, then the

observed effect cannot be said to be due to unmeasured confounding. This method is very simple to apply, and can be applied when information on the relationship between exposure and confounder is available. When this information is not available, Bross[123] provided an extension to allow the calculation of the minimum value of the association between the unobserved confounder and outcome that could explain the observed exposure-outcome association for various relationships between the exposure and confounder. This method only applies to binary outcome and exposure and a single binary confounder.

Schlesselman[124] extended Bross's[122] method to allow for an interaction between the exposure and confounder. In the method, the observed exposure-outcome relative risk is adjusted for an unmeasured confounder using the relative risk of disease due to the confounder in the absence of the exposure, the relative risk of disease due to the confounder in the presence of the exposure, and the prevalences of the confounder in the exposed and unexposed populations. It was shown that, if there is an interaction between the exposure and confounder, previously described methods[121, 122] may either underestimate or overestimate the effect of the unmeasured confounder. Once again, it was assumed that the outcome and exposure were binary, and that there was a single unmeasured binary confounder. The method used by Schlesselman to estimate the prevalences of the confounder in the exposed and unexposed populations has been corrected by Simon,[125] who also noted that correcting the observed confidence limits in the same way as the relative risk (by dividing by the effect of the unmeasured confounder on the estimated exposure-outcome association) is only appropriate in large samples.

A method for assessing the sensitivity of results to an unmeasured binary confounder was described by Rosenbaum and Rubin.[126] The method assumes a binary outcome, binary exposure, and a categorical confounder. The sensitivity parameters are the log ORs between the exposure and the unmeasured confounders, the log OR between the outcome and the unobserved confounder, and the probability that the unobserved confounder is zero. These sensitivity parameters can vary across strata of the observed confounder, although the sensitivity analysis could become very large if all parameters are allowed to vary across all categories of the observed confounder. Estimates of the expectations of the proportion of subjects improved by exposure can then be calculated using various values of the sensitivity parameters.

Yanagawa[127] presented a method for calculating the upper and lower bounds on an estimated exposure-outcome OR from a case-control study in which there is unmeasured confounding. It was assumed that disease, exposure and the unmeasured confounder were binary variables. The adjustment used the confounder-outcome ORs among the exposed and unexposed subjects, and the exposure-confounder ORs among the cases and controls. One problem with this type of sensitivity analysis is that the sensitivity parameters are treated as known. Yanagawa[127] dealt

with this complication by suggesting that the 95% confidence intervals for the sensitivity parameters obtained from external studies be used to calculate conservative bounds on the estimated exposure-outcome OR.

Lin, Psaty and Kronmal[128] presented a method for assessing the sensitivity of results from regressions on binary or censored failure time outcomes to unmeasured confounding. For binary outcomes, the model described is slightly different from the logistic regression model, but is a good approximation of it for rare outcome events. For censored failure time outcomes, a Cox proportional hazards model is used. When the unmeasured confounder is binary, the sensitivity parameters are the association between the confounder and outcome in the exposed and unexposed groups, and the prevalences of the confounder in the exposed and unexposed groups. If the unmeasured confounder is normally distributed, the sensitivity parameters are the associations between the confounder and outcome in the exposed and unexposed groups, and the means of the confounder in the exposed and unexposed groups. These sensitivity parameters are then used to adjust the observed exposure-outcome association. This method can only assess the impact of a single unmeasured confounder.

A method to investigate the impact of several unmeasured confounders has been proposed by Stürmer, Schneeweiss, Avorn, *et al.*[129] This propensity score calibration method combines propensity scores with regression calibration. Propensity scores, which are the conditional probability of exposure given the other covariates, are estimated in the main study in which information on all relevant confounders is not available. A validation study, in which information on additional confounders is available, is used to estimate the relationship between the propensity score which is subject to unmeasured confounding and the propensity score in which all relevant confounders are included. The main study estimates, in which the propensity scores are subject to unmeasured confounding, are then corrected using the estimated relationship between the confounded and unconfounded propensity scores from the validation data.

4.4.2. External adjustment

A simple method of sensitivity analysis, given by Greenland,[60, 130, 131] is the method of external adjustment, which can assess the impact of unmeasured confounding, misclassification, and selection bias. External adjustment involves using sources of data external to the study in question to adjust the observed effect estimates. Unmeasured confounding is adjusted for by assuming values for the association between outcome and the unmeasured confounder, and prevalences of the confounder in the exposed and unexposed populations. The impact of misclassification of a dichotomous exposure variable is investigated by assuming different values for the sensitivity and specificity of misclassification. Selection bias is analysed by assuming probabilities of selection. For all of the above corrections, assuming different values

for the required parameters leads to a sensitivity analysis. This method of sensitivity analysis could lead to a large number of parameters being varied, particularly if more than one source of bias or variable are being investigated. Presentation of results from such a sensitivity analysis would be complex, and possibly difficult to interpret. The methods apply to categorical variables, and were illustrated using the special case of binary variables, but are not applicable to continuously distributed variables.

Gail, Wacholder and Lubin[132] described a method of external adjustment to assess the sensitivity of estimated relative risks to a binary unmeasured confounder. Adjustments to the observed crude relative risk estimate were made using the confounder-outcome relative risk and the probability of having the confounder in the exposed and unexposed populations. External adjustment for additive models involves correcting the observed risk difference using the confounder-exposure risk difference, and again the probability of having the confounder in the exposed and unexposed groups.

4.4.3. Monte Carlo risk analysis

Greenland[133] used the term "Monte Carlo risk analysis" to refer to methods in which sensitivity parameters are drawn from a probability distribution, and the conventional analysis is repeated for multiple draws.

A Monte Carlo risk analysis method was described by Lash and Fink,[61] in which the counterfactual dataset was reconstructed. A counterfactual dataset is the dataset that would have been observed had the various sources of bias, such as misclassification, selection bias and unmeasured confounding not been present. Selection bias in the example given was due to outcome status being unknown for some of the study subjects. This was adjusted for by "guessing" the status of these subjects, informed by the data for subjects whose outcome was known.

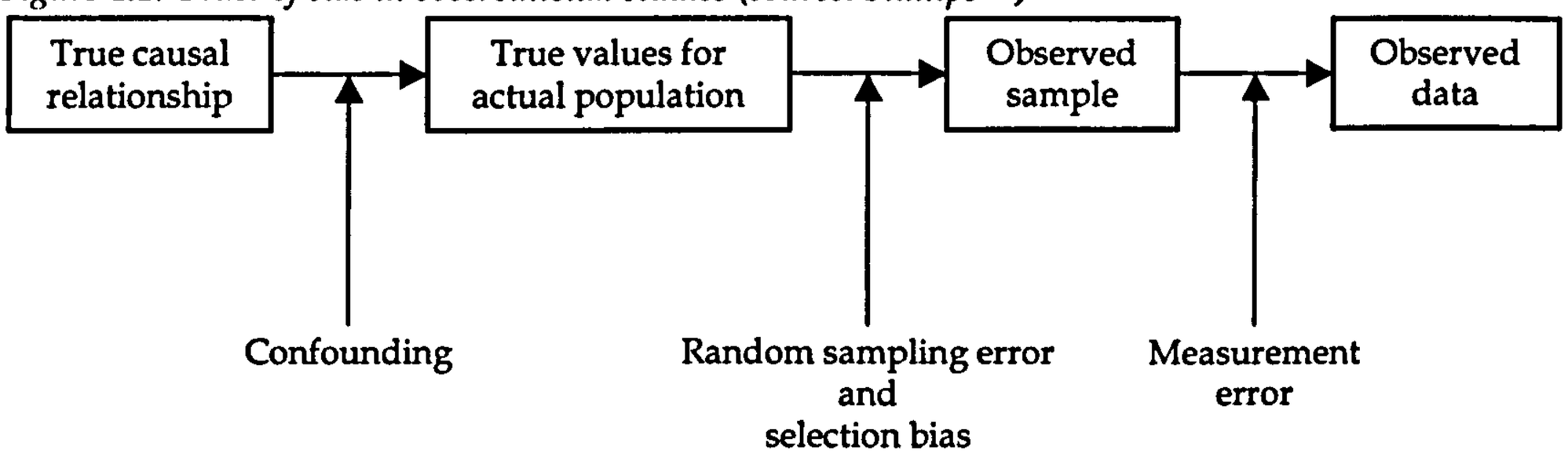
To assess the impact of misclassification in a binary explanatory variable, triangular probability distributions were defined to represent the sensitivity and specificity of the variable. For a single reconstruction, a sensitivity and specificity were chosen from the probability distributions and used to calculate the positive and negative predictive values respectively. For each subject who may have been misclassified, a Bernoulli trial was performed using either the positive or negative predictive value, depending on the subject's observed status, to model whether misclassification had occurred. Subjects identified as misclassified were then reclassified. A SAS macro to investigate the impact of misclassification of binary variables on estimated exposure-outcome associations was developed by Fox, Lash and Greenland.[134] The macro allows for uniform, triangular and trapezoidal probability density functions.

To assess the effect of an unmeasured confounder, Lash and Fink[61] created a dichotomous variable by specifying the prevalence of the confounder in subsets of the population. A Bernoulli trial was performed to assign whether a subject had the confounder.

Once the counterfactual dataset had been reconstructed, standard analyses were used to obtain an estimate of the exposure-outcome association. Multiple reconstructions were made, and the 50th percentile of the cumulative probability distribution was used to define an overall point estimate. The 2.5th and 97.5th percentiles were used as the 95% simulation interval. This analysis ignores sampling variation (the error due to analysing a sample of the true population), so a bootstrap sample of each reconstructed dataset was taken, and the same method used to obtain a point estimate and 95% simulation interval.

A similar method, with the same data as used by Lash and Fink,[61] appeared in Lash and Silliman.[135] Phillips[136] also presented a similar method, although the order in which the corrections should be applied was emphasized. Figure 4.1 shows the order in which bias in observational studies occurs, from left to right. Phillips argued that the corrections in the sensitivity analysis should be performed from right to left.

Figure 4.1: Order of bias in observational studies (source: Phillips[136])



4.4.4. Other methods

Sensitivity analysis methods for specific analysis models have been proposed.

Test statistics are often used in observational epidemiological studies to test for relationships between exposure and outcome. In these analyses, it is assumed that the exposure in each pair is randomly assigned. Rosenbaum[137-140] and Rosenbaum and Krieger[141] have presented a sensitivity analysis that investigates the impact of degrees of departure from this assumption of random assignment of exposure. Departure from the assumption is parameterised using a sensitivity parameter. This parameter is then varied, and bounds on the significance level of the test statistic calculated to show how conclusions could be altered with different amounts of non-random treatment assignment. This approach can be applied to many test statistics, such as Wilcoxon's signed rank test for matched pairs with continuous outcomes, the McNemar-Cox

test for paired binary outcomes, Gehan and log-rank tests for censored outcomes, and Fisher's exact test for binary responses.

Jo[142] proposed a sensitivity analysis method to investigate the impact of model misspecification when estimating treatment efficacy in RCTs in which there are departures from randomly allocated treatment. One method of estimating treatment efficacy in RCTs is to estimate the complier average causal effect (CACE),[143] which is described in more detail in Section 8.2.1.1. CACE estimation relies on several assumptions including the exclusion restriction assumption. This assumption states that, if randomisation does not affect the treatment actually received, then randomisation also has no effect on outcome. The proposed sensitivity analysis investigates the impact of violation of the exclusion restriction assumption in CACE estimation.

4.4.5. Summary

Only a single source of bias is considered in the sensitivity analysis methods for the impact of unmeasured confounding on exposure effect estimates. Epidemiological studies may be subject to several sources of bias, such as measurement error or selection bias, and therefore using a sensitivity analysis to allow for just unmeasured confounding may not be adequate. Sensitivity analysis methods that allow for multiple sources of bias simultaneously, such as external adjustment and Monte Carlo risk analysis, may therefore be preferable to methods that allow for only a single source of bias. Presentation of the results of a sensitivity analysis may also cause problems. If several sources of bias are being investigated, and several values of the relevant parameters are used for each source of bias, a large number of results will be obtained, and presenting these in a comprehensible way will be difficult. This problem affects all of the sensitivity analysis methods described in this section, other than Monte Carlo risk analysis.

The motivation for the literature described in this section is an analysis of the effect of residual confounding (caused by measurement error in the confounders) and exposure measurement error on the association between C-reactive protein and coronary heart disease, which will be presented in Chapter 7. The data for this analysis do not allow estimation of the measurement error parameters, and therefore sensitivity analysis methods must be used. The impact of unmeasured confounding will not be considered in Chapter 7, and therefore the sensitivity analysis methods for unmeasured confounding described in Section 4.4.1 above will not be used. In common with the simulation studies presented in Chapter 3, only measurement error in continuous variables will be considered. External adjustment (Section 4.4.2) will therefore also not be used. Monte Carlo risk analysis will be used. An advantage of this method over many other sensitivity analysis methods is that the results are easy to display in terms of a point estimate and simulation interval. Monte Carlo risk analysis is also applicable to a range of regression models, is able to investigate the joint impact of unmeasured confounding, measurement error in explanatory variables and selection bias on exposure effect estimates, and

is applicable to both categorical and continuous explanatory variables. This makes it an attractive option for any sensitivity analysis.

BLANK IN ORIGINAL

Chapter 5.

Correcting for measurement error in explanatory variables in epidemiological studies: A simulation study

5.1. Introduction

In this chapter, two methods for correcting for measurement error, simulation-extrapolation (SimEx) and regression calibration, are investigated using simulation studies. Correction for residual confounding, caused by measurement error in the confounders, is investigated in simulated datasets which contain either two or four confounders. For datasets with two confounders, each may be measured with error or omitted from the analysis. For datasets with four confounders, two confounders are omitted from every analysis, and the remaining two may be either measured with error or omitted from the analysis. Correction for exposure measurement error, as well as residual confounding, is investigated using simulated datasets with two confounders, where again the confounders may be measured with error or omitted from the analysis. For all analyses, logistic regression is used to estimate the corrected exposure-outcome odds ratios (ORs) in situations in which the exposure has no true association with the outcome (exposure-outcome OR=1).

5.2. Methods

5.2.1. Notation and terminology

The following is a list of the notation that will be used throughout this chapter.

σ_u^2 = variance of the true measurements

σ_e^2 = measurement error variance

μ_s = mean of the observed measurements of S

W = vector of variables observed without error

X = vector of unobserved perfectly measured variables

Z = vector of observed variables measured with error

Σ_{st} = variance-covariance matrix between variables S and T

$\hat{X}$ = estimated vector of perfectly measured variables

k = number of replicates of the observed variables measured with error

Throughout this chapter, the results presented in Chapter 3 will be referred to as *naïve estimates,* and corresponding analysis method as the *naïve analysis.* Results corrected by the SimEx method will be referred to as the *SimEx corrected estimates,* and results corrected by the regression calibration method will be referred to as the *regression calibration corrected estimates.* The term *overcorrection* will refer to corrected estimates that are further from the naïve estimates than the estimate obtained when the variables included in the analysis are measured without error. This is not necessarily the true exposure-outcome OR, as the estimate may be biased by unmeasured confounding. The term *intermediate* will refer to corrected estimates that are between the naïve estimate and the estimate obtained when there is no measurement error in

any of the variables included in the analysis. The corrected estimates will generally be compared with the estimates obtained when there is no measurement error in the confounders included in the analysis as, although the aim is to recover the true exposure-outcome OR, in the presence of unmeasured confounding this will be impossible using methods to correct for measurement error. When comparing two estimates, the term *more extreme* will be used to refer to the fact that one estimate is further from the true exposure-outcome OR than the other, regardless of the direction of the estimated effect.

Due to the complexity of the tables presented in this chapter, where examples are given in the text the relevant results in the tables will be labelled.

5.2.2. Simulated datasets

The datasets were simulated in the same way as described in Section 3.2.2. Briefly, the datasets consisted of an exposure and either two or four confounders, all with standard normal distributions. The exposure-outcome log OR was zero, while the confounder-outcome log ORs were ln2 per standard deviation increase for all confounders. Datasets in which the exposure was causally related to the outcome were not simulated. The correlations between the exposure and each confounder was 0.1, 0.3 or 0.5, and the correlations between pairs of confounders was either 0 or 0.5. If four confounders were simulated, the correlations between pairs of confounders were equal for all pairs. Measurement error was introduced into the exposure and each confounder to generate intra-class correlation coefficients (ICCs) of either 0.75 or 0.5, which correspond to error variances of 1/3 or 1 respectively. Justification for these choices of correlations and ICCs was provided in Section 3.2.2. For simulations in which the exposure was measured with error, the error-prone exposure variable was transformed to have variance equal to one. The estimated exposure-outcome ORs can then be interpreted as the OR per standard deviation increase in the error-prone exposure variable. The binary outcome was generated by simulating a Bernoulli trial for each observation in the dataset, using the probability of outcome defined by the logistic regression model. For each simulated dataset, 500,000 observations were generated. Following simulation of the datasets, SimEx or regression calibration was used to correct the estimates for exposure measurement error and residual confounding. This simulation and correction process was repeated 50 times for each combination of correlations between explanatory variables. The ORs presented in this chapter are the geometric means of the corrected ORs from each of the 50 repetitions. Fifty simulations each of 500,000 observations resulted in small 95% simulation intervals around the geometric means, with widths no more than 0.014.

5.2.3. Simulation-extrapolation

The SimEx method was originally proposed by Cook and Stefanski.[81] First, the observed data is used to estimate the uncorrected parameter estimates, and if necessary the measurement error

variances. Pseudoerrors are then generated to be mutually independent, independent of the observed data and the true underlying variables, and identically distributed standard normal variables. These pseudoerrors are used to add extra measurement error to the mismeasured variables. The amount of measurement error added is defined by a scale factor λ and the estimated or known measurement error variance. A scale factor of zero corresponds to using the observed variables with no change. Scale factors greater than zero correspond to increasing amounts of measurement error. The variables with extra measurement error are then used in the required regression analysis to obtain an estimate of the parameter of interest.

The parameter estimates for increasing amounts of measurement error can then be used to extrapolate back to the parameter that would be estimated if there was no measurement error. This corresponds to λ=-1. There are many forms that the extrapolant function can take. Hardin, Schmiediche and Carroll[144] have written a Stata command, simex, which implements the SimEx method for generalised linear models with measurement error. The command allows use of quadratic, rational linear, and simple linear extrapolants. The quadratic extrapolant involves fitting the following model to the data, where $\hat{\theta}$ is the estimated parameter of interest from the error-prone data.

$$\hat{\theta} = \beta_0 + \beta_1\lambda + \beta_2\lambda^2 .$$

The rational linear extrapolant takes the following form:

$$\hat{\theta} = \beta_0 + \frac{\beta_1}{\beta_2 + \lambda},$$

and the simple linear extrapolant is the following:

$$\hat{\theta} = \beta_0 + \beta_1\lambda .$$

The corrections in this chapter used the quadratic extrapolant.

For each value of the scale factor, the pseudoerrors are generated B times, and the estimate of the parameter of interest is taken to be the mean of the parameters obtained from each of the B repetitions. In this chapter, B was equal to 50. This process is repeated for increasing values of the scale factor.

The SimEx method for the datasets containing 500,000 observations is computationally intensive. Simulations are therefore not performed for correlations between pairs of confounders of 0.1, 0.2, 0.3 or 0.4. In addition, for simulations with four confounders, the correlation between exposure and X_3 is fixed to be 0.5, and the correlation between exposure and X_4 is set to be 0.3. These values correspond to the values used for the results presented in Chapter 3.

5.2.3.1 Example analysis

Table 5.1: Example data from a simulation

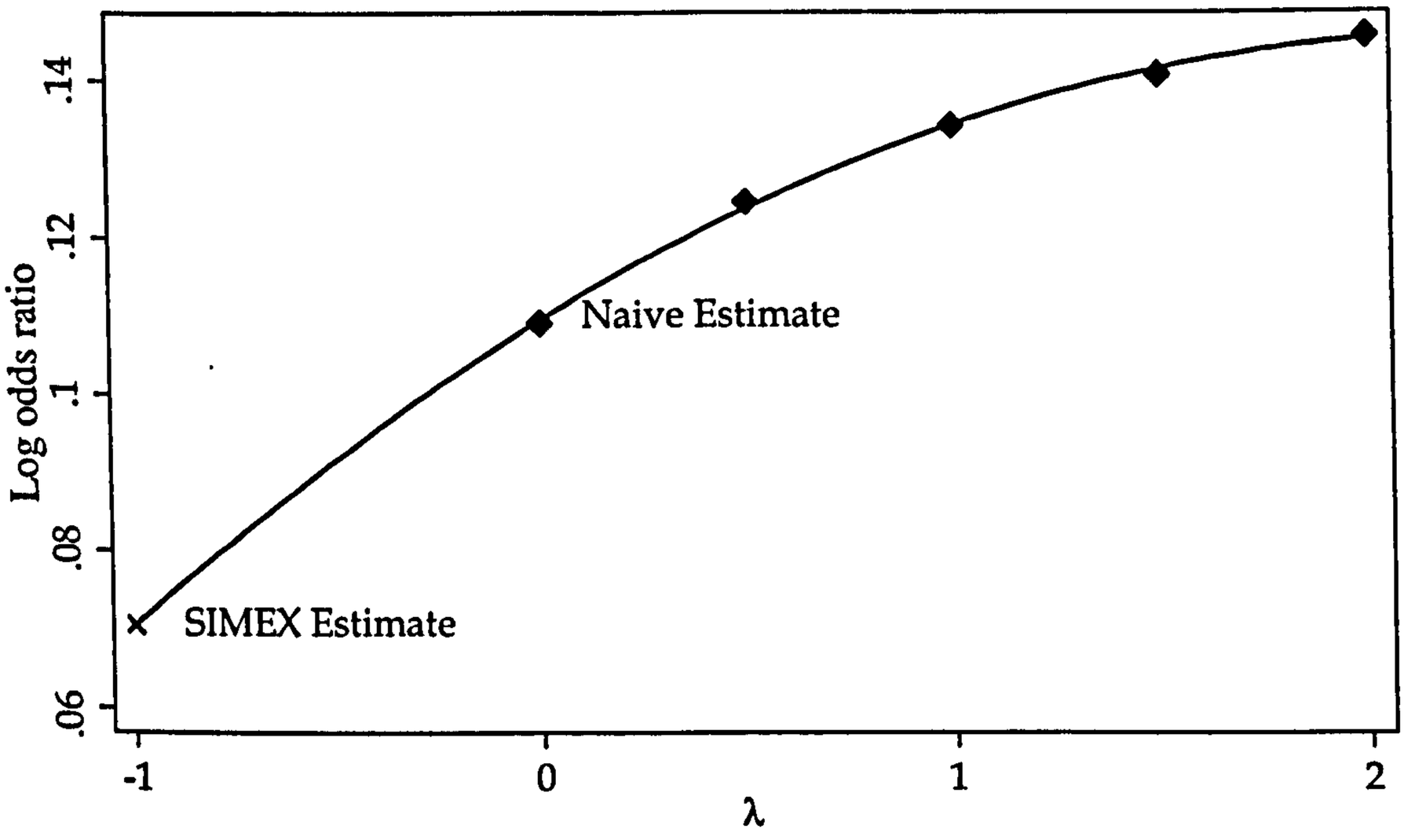
Subject id	Outcome	Exposure	Confounder 1 (ICC=1)	Confounder 2 (ICC=0.5)	Confounder 2 (Estimated)[a]
1	0	-1.85	-0.0039	1.21	0.21
2	0	0.44	-0.049	-1.13	-0.41
3	0	2.12	-1.64	-1.08	-0.58
4	1	0.33	0.50	1.21	0.68
5	0	-0.69	-1.30	-1.44	-1.05
6	0	1.27	0.63	0.55	0.59
7	0	-0.43	-1.02	0.58	-0.11
8	0	-1.75	-0.062	-0.048	-0.30
9	0	-0.49	0.26	-0.84	-0.34
10	1	0.77	-0.24	1.11	0.50
11	0	1.61	0.84	1.91	1.25
12	0	-1.37	-0.16	2.19	0.64
13	1	1.26	0.31	1.90	1.05
14	0	0.69	-1.89	-2.02	-1.25
15	0	0.79	1.06	-0.22	0.33
16	0	-1.07	-0.57	-0.76	-0.63
17	0	1.47	1.90	-0.29	0.64
18	0	-1.97	0.53	0.22	-0.057
19	0	1.81	-0.11	0.27	0.35
20	0	-0.55	-0.56	0.59	0.00014

[a] *This column does not relate to the example of SimEx correction, but shows the estimated values of the perfectly measured confounder 2 from the example analysis using regression calibration correction described in Section 5.2.4.1.*

Table 5.1 shows the first 20 observations from a simulated dataset of 500,000 observations with two confounders in which confounder 2 is measured with error, with an ICC of 0.5. The exposure and confounders are all standard normal variables. The correlation between the exposure and confounder 1 is 0.1, the correlation between exposure and confounder 2 (measured without error) is 0.3, and the correlation between the two confounders (measured without error) is 0.5. The true exposure-outcome OR is one, and the ORs between the perfectly measured confounders and outcome are both two. A logistic regression analysis to estimate the association between exposure and outcome, adjusting for confounder 1 and the error-prone measurements of confounder 2, gives an estimated exposure-outcome OR of 1.12.

Figure 5.1 shows the simulation-extrapolation results for this example dataset. The naïve point estimate is plotted at the value λ=0. Measurement error is then added to confounder 2, and the point estimate plotted at values of λ greater than one. These values are then extrapolated back to the value λ=-1 to provide the SimEx estimate of the exposure-outcome OR of 1.07.

Figure 5.1: Results of simulation extrapolation for an example dataset.



Naïve estimated odds ratio: 1.12 SIMEX corrected odds ratio: 1.07

5.2.4. Regression calibration

The regression calibration method used in this chapter is the one described by Carroll and Stefanski.[70] The basis of this method is to estimate the values of the unobserved true covariates, based on the values of the mismeasured covariate and any perfectly measured covariates. The estimated values of the unobserved variables are then used in a regression analysis to provide an estimate that is corrected for measurement error.

The unobserved variables, $\mathbf{X}$, are estimated by the following equation:

Equation 5.1: $$\hat{\mathbf{X}} \approx \mu_X + \begin{pmatrix} \Sigma_{xx} \\ \Sigma_{wx} \end{pmatrix}^T \begin{pmatrix} \Sigma_{xx} + \Sigma_{uu}/k & \Sigma_{xw} \\ \Sigma_{wx} & \Sigma_{ww} \end{pmatrix}^{-1} \begin{pmatrix} \overline{Z}_{i\bullet} - \mu_Z \\ W - \mu_W \end{pmatrix}$$

where W denotes the observed error-free variables, $\overline{Z}_{i\bullet}$ is the mean of the repeated measurements of the error-prone observed variables, μ_X is the mean of variable X, Σ_{xw} is the variance-covariance matrix between variables X and W, Σ_{uu} is the covariance matrix of the measurement errors, and k is the number of repeated measurements of the error-prone variables. If there is only one measurement of an error-prone variable available in the dataset, $k=1$ and $\overline{Z}_{i\bullet}$ is replaced by Z. For the analysis in this chapter, the error variance was known, and there was a single replicate of each error-prone variable. Note that not all parameters in Equation 5.1 are observed, because X is unobserved. The variance-covariance matrices are replaced by matrices estimated using the observed variables. The mean of the unobserved variables, μ_X, is replaced by $\hat{\mu}_Z$, because the measurement error is assumed to be random, and therefore the expected values of the measurement errors is zero. Calculation of the estimated covariance matrices is described below for the situation in which there is a single measurement

of the error-prone variables and the covariance matrix of the measurement errors is known.

The estimated variance-covariance matrix between the observed perfectly measured variables is given by

$$\hat{\Sigma}_{ww} = \frac{1}{n-1}\sum_{i=1}^{n}(W_i - \overline{W})(W_i - \overline{W})^T,$$

the estimated variance-covariance matrix between the unobserved and observed perfectly measured variables is given by

$$\hat{\Sigma}_{xw} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z})(W_i - \overline{W})^T,$$

and the estimated variance-covariance matrix of the unobserved perfectly measured variables is given by

$$\hat{\Sigma}_{xx} = \frac{1}{n-1}\left[\sum_{i=1}^{n}(Z_i - \overline{Z})(Z_i - \overline{Z})^T\right] - \Sigma_{uu}.$$

The unobserved perfectly measured variables are therefore estimated using the following equation:

$$\hat{X} \approx \hat{\mu}_Z + \begin{pmatrix}\hat{\Sigma}_{xx}\\ \hat{\Sigma}_{wx}\end{pmatrix}^T \begin{pmatrix}\hat{\Sigma}_{xx}+\Sigma_{uu} & \hat{\Sigma}_{xw}\\ \hat{\Sigma}_{wx} & \hat{\Sigma}_{ww}\end{pmatrix}^{-1} \begin{pmatrix}Z_i - \hat{\mu}_Z\\ W_i - \hat{\mu}_W\end{pmatrix}.$$

These equations are similar to those given in Hardin, Schmiediche and Carroll[145] and Carroll, Ruppert and Stefanski,[65] with differences due to notation and the fact that in the analysis presented in this chapter, k=1 and the measurement error variance-covariance matrix is known.

The Stata command rcal, written by Hardin, Schmiediche and Carroll,[145] is used to correct the simulation studies for residual confounding.

5.2.4.1 Example analysis

The same dataset is used for this example analysis as for the example SimEx analysis described in Section 5.2.3.1. Table 5.1 shows the first 20 observations of the simulated dataset. A logistic regression analysis to estimate the association between exposure and outcome, adjusting for confounder 1 and the error-prone measurements of confounder 2, gives an estimated exposure-outcome OR of 1.12. The final column in Table 5.1 shows the values of confounder 2 estimated by the regression calibration procedure. To estimate these values, the covariance matrices used were

$$\Sigma_{uu} = (1),$$

$$\hat{\Sigma}_{ww} = \begin{pmatrix}1 & 0.1\\ 0.1 & 1\end{pmatrix},$$

$$\hat{\Sigma}_{xw} = (0.3 \quad 0.5), \text{ and}$$

$$\hat{\Sigma}_{xx} = (1).$$

A logistic regression analysis to estimate the association between exposure and outcome, adjusting for confounder 1 and using the estimated values of confounder 2 rather than the error-prone measurements, gives an estimated exposure-outcome OR of 1.01.

5.3. Results

5.3.1. Two confounders

In this section, the results for simulations with two confounders, which may be measured with error or omitted from the regression analysis, are presented. Throughout this section the exposure is assumed to be measured without error.

5.3.1.1 Simulation-extrapolation

Table 5.2 shows the results of SimEx correction for the case in which the confounders are uncorrelated. The bold numbers are the SimEx corrected estimates, while the numbers in italics are the naïve estimates (as shown in Table 3.1). For analyses in which the confounder or confounders were not measured with error, the SimEx corrected estimate was not calculated. This is indicated in Table 5.2 by dashes.

The SimEx corrected estimate is always less biased than the naïve estimate. However, the SimEx method very rarely removes all of the bias due to residual confounding. Out of the 90 pairs of naïve and SimEx corrected estimates presented in Table 5.2, only two of the SimEx corrected estimates have removed all bias. It is not surprising that in analyses in which only Z_1 was controlled, the SimEx method does not remove all bias. In these analyses, the bias is due to both residual and unmeasured confounding. The SimEx method only attempts to correct for the confounder measurement error that results in residual confounding of the exposure-outcome OR.

Considering the SimEx corrected estimates only, bias is increased by unmeasured confounding, more measurement error (lower ICC) in the confounders, and by greater correlation between the exposure and the perfectly measured confounders. These patterns are the same as those observed for the naïve estimates presented in Table 3.1.

Table 5.3 shows the results of SimEx correction when the confounders have a correlation of 0.5. Again, the bold numbers are the SimEx corrected estimates, the numbers in italics are the naïve estimates (as shown in Table 3.2), and a dash indicates analyses in which SimEx correction was not used.

The SimEx corrected estimates are generally less biased than the naïve estimates. There is, however, one exception to this rule. Consider when the correlation between E and X_1 is 0.5, the

Table 5.2: Geometric means of the simulation-extrapolation corrected estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, sample size=500,000, repetitions=50. Numbers in bold are the simulation-extrapolation corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1				0.3				0.5			
		OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
			ICC of Z_2				ICC of Z_2				ICC of Z_2		
			0.5	0.75	1		0.5	0.75	1		0.5	0.75	1
0.1	0.5	**1.09**	**1.04**	**1.02**	**1.02**	**1.24**	**1.08**	**1.04**	**1.02**	**1.42**	**1.15**	**1.06**	**1.03**
		1.10	*1.07*	*1.05*	*1.04*	*1.25*	*1.15*	*1.10*	*1.04*	1.43	*1.26*	*1.16*	*1.05*
	0.75	**1.07**	**1.02**	**1.01**	**1.00**	**1.23**	**1.07**	**1.02**	**1.01**	**1.40**	**1.13**	**1.04**	**1.01**
		1.09	*1.05*	*1.04*	*1.02*	1.24	*1.13*	*1.08*	*1.02*	*1.42*	*1.24*	*1.14*	*1.02*
	1	-	**1.02**	**1.00**	-	-	**1.06**	**1.02**	-	-	**1.13**	**1.03**	-
		1.07	*1.04*	*1.02*	*1.00*	*1.22*	*1.11*	*1.06*	*1.00*	*1.40*	*1.22*	*1.11*	*1.00*
0.3	0.5	**1.13**	**1.08**	**1.07**	**1.06**	**1.30**	**1.14**	**1.09**	**1.07**	**1.51**	**1.23**	**1.13**	**1.09**
		1.18	*1.15*	*1.13*	*1.11*	*1.36*	*1.24*	*1.18*	*1.12*	*1.56*	*1.38*	*1.27*	*1.16*
	0.75	**1.09**	**1.04**	**1.02**	**1.02**	**1.26**	**1.09**	**1.03**	**1.02**	**1.46**	**1.17**	**1.06**	**1.02**
		1.13	*1.10*	*1.08*	*1.06*	*1.30*	*1.18*	*1.13*	*1.06*	*1.50*	*1.31*	*1.20*	*1.08*
	1	-	**1.02**	**1.01**	-	-	**1.07**	**1.02**	-	-	**1.14**	**1.04**	-
		1.08	*1.04*	*1.02*	*1.00*	*1.24*	*1.12*	*1.06*	*1.00*	*1.44*	*1.24*	*1.13*	*1.00*
0.5	0.5	**1.21**	**1.15**	**1.13**	**1.13**	**1.42**	**1.23**	**1.17**	**1.14**	**1.68**	**1.37**	**1.25**	**1.19**
		1.29	*1.26*	*1.24*	*1.22*	*1.51*	*1.38*	*1.31*	*1.24*	*1.76*	*1.56*	*1.44*	*1.31*
	0.75	**1.12**	**1.06**	**1.04**	**1.03**	**1.33**	**1.13**	**1.06**	**1.04**	**1.59**	**1.25**	**1.12**	**1.05**
		1.20	*1.16*	*1.14*	*1.11*	*1.41*	*1.27*	*1.20*	*1.13*	*1.67*	*1.44*	*1.31*	*1.17*
	1	-	**1.03**	**1.01**	-	-	**1.09**	**1.02**	-	-	**1.19**	**1.05**	-
		1.09	*1.05*	*1.02*	*1.00*	*1.30*	*1.16*	*1.08*	*1.00*	*1.56*	*1.31*	*1.17*	*1.00*

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0056.*

Table 5.3: Geometric means of the simulation-extrapolation corrected estimated exposure-outcome odds ratio for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, sample size=500,000, repetitions=50. Numbers in bold are the simulation-extrapolation corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 — OR adjusted for Z_1	0.1 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.75	0.1 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 1	0.3 — OR adjusted for Z_1	0.3 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.3 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.75	0.3 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 1	0.5 — OR adjusted for Z_1	0.5 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.5 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.75	0.5 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 1
0.1	0.5	**1.06** *1.08*	**1.02** *1.05*	**1.01** *1.04*	**1.01** *1.02*	**1.21** *1.23*	**1.05** *1.12*	**1.00** *1.05*	**0.99** *0.98*	**1.38** *1.39*	**1.10** *1.20*	**0.99** *1.07*	**0.95** *0.93*
	0.75	**1.04** *1.06*	**1.01** *1.04*	**1.00** *1.03*	**1.00** *1.01*	**1.19** *1.21*	**1.06** *1.11*	**1.01** *1.05*	**1.00** *0.99*	**1.36** *1.38*	**1.12** *1.21*	**1.02** *1.10*	**0.99** *0.96*
	1	- *1.03*	**1.01** *1.02*	**1.00** *1.01*	- *1.00*	- *1.18*	**1.07** *1.11*	**1.02** *1.06*	- *1.00*	- *1.36*	**1.15** *1.22*	**1.05** *1.13*	- *1.00*
0.3	0.5	**1.05** *1.12*	**1.05** *1.12*	**1.06** *1.11*	**1.07** *1.11*	**1.21** *1.28*	**1.08** *1.18*	**1.05** *1.13*	**1.04** *1.07*	**1.39** *1.46*	**1.13** *1.27*	**1.04** *1.16*	**1.02** *1.03*
	0.75	**0.99** *1.04*	**1.00** *1.05*	**1.01** *1.05*	**1.02** *1.06*	**1.14** *1.20*	**1.05** *1.13*	**1.02** *1.09*	**1.01** *1.04*	**1.32** *1.38*	**1.11** *1.23*	**1.03** *1.13*	**1.01** *1.02*
	1	- *0.97*	**0.99** *0.98*	**1.00** *0.99*	- *1.00*	- *1.12*	**1.04** *1.07*	**1.01** *1.04*	- *1.00*	- *1.29*	**1.11** *1.18*	**1.03** *1.10*	- *1.00*
0.5	0.5	**1.05** *1.17*	**1.10** *1.20*	**1.12** *1.21*	**1.15** *1.22*	**1.23** *1.36*	**1.13** *1.27*	**1.11** *1.23*	**1.11** *1.18*	**1.44** *1.57*	**1.19** *1.39*	**1.11** *1.27*	**1.09** *1.15*
	0.75	**0.92**[b] *1.03*[a]	**0.99**[e] *1.07*[d]	**1.02** *1.10*	**1.05** *1.13*	**1.09** *1.21*	**1.04** *1.16*	**1.03** *1.13*	**1.03** *1.10*	**1.30** *1.42*	**1.11** *1.27*	**1.04** *1.18*	**1.03** *1.08*
	1	- *0.88*[c]	**0.95** *0.93*	**0.99** *0.96*	- *1.00*	- *1.05*	**1.02** *1.03*	**1.01** *1.02*	- *1.00*	- *1.25*	**1.09** *1.15*	**1.03** *1.08*	- *1.00*

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_U=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0046.*

[a, b, c, d, e] *These results are referred to in the text.*

correlation between E and X_2 is 0.1, and the ICC of Z_1 is 0.75. In the naïve analysis, the OR when adjusting for Z_1 only is 1.03 (labelled a in Table 5.3). The SimEx corrected estimate, however, is 0.92 (labelled b in Table 5.3). This corrected estimate is more extreme than the naïve estimate, and the effect is in the opposite direction. This is, however, not an overcorrection by the SimEx method, because the corrected estimate of 0.92 is intermediate between the naïve estimate of 1.03 and the estimated OR of 0.88 (labelled c in Table 5.3) which is not subject to residual confounding by X_1 but is biased by unmeasured confounding by X_2.

Generally, the SimEx corrected estimate is intermediate between the naïve estimate and the estimate that would be obtained in the absence of residual confounding. There is one exception to this rule, where the SimEx method overcorrects the naïve estimate. Consider when the correlation between E and X_1 is 0.5, and the correlation between E and X_2 is 0.1. For analyses in which both confounders are controlled for, when the ICC of Z_1 is 0.75 and the ICC of Z_2 is 0.5 the naïve estimate equals 1.07 (labelled d in Table 5.3). The SimEx corrected estimate in this situation is 0.99 (labelled e in Table 5.3), which is further from the naïve estimate than the true exposure-outcome OR of 1.00. This overcorrection is, however, very small.

The SimEx method seems to be slightly better at correcting for measurement error when the confounders are correlated than when the confounders are uncorrelated. In Table 5.3, seven of the SimEx corrected estimates have completely removed the bias observed in the naïve estimate.

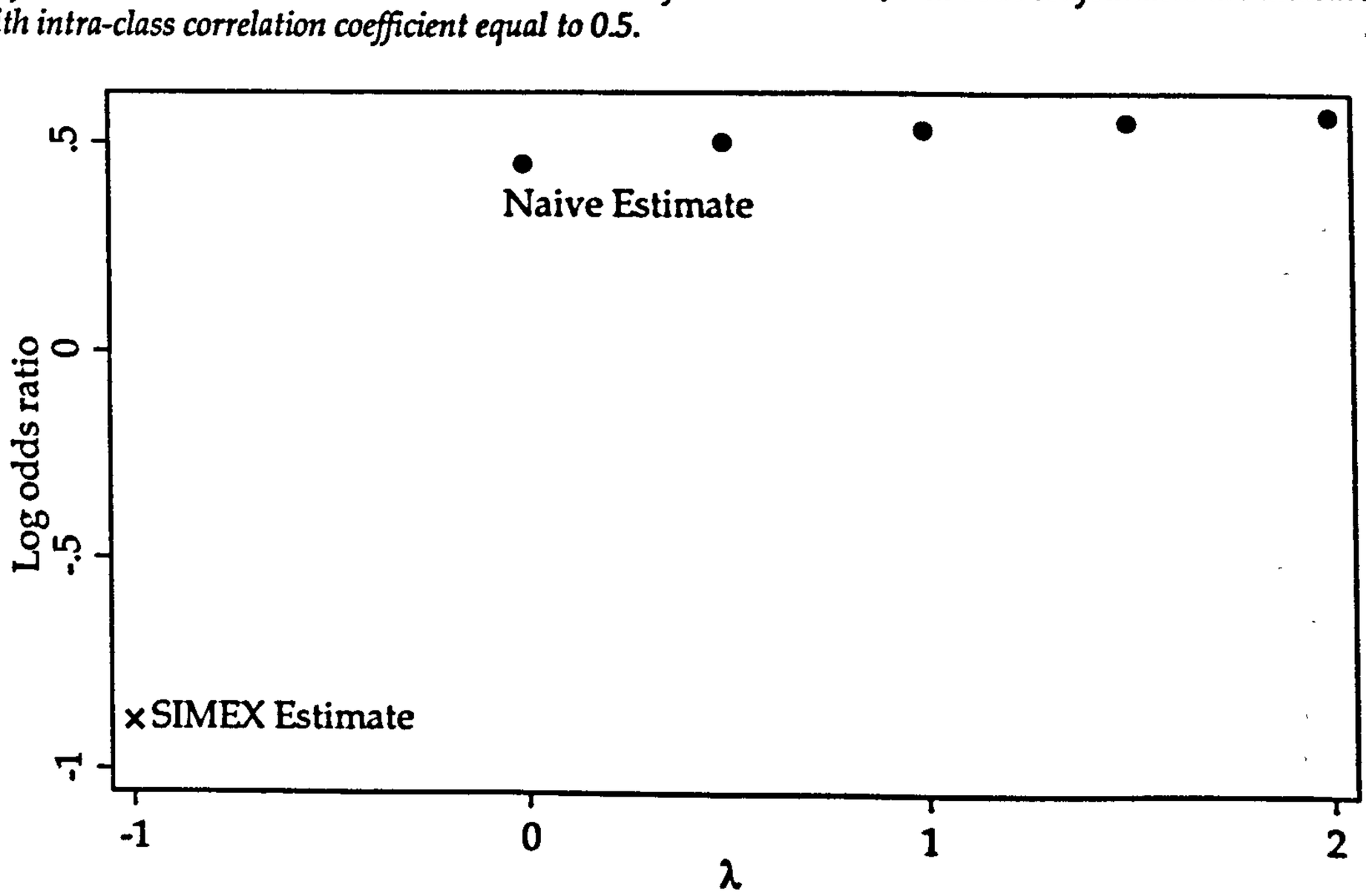
Considering the SimEx corrected estimates only, in general bias in the estimates increases with unmeasured confounding, increasing measurement error in the confounders (decreasing ICC), and increasing correlation between E and the perfectly measured confounders. There are, however, exceptions to all of these generalisations, as was also the case with the results presented in Table 3.2. The bias in the SimEx corrected estimates may decrease as unmeasured confounding, measurement error in the confounders, or the correlation between the confounders and exposure increases.

Correcting for measurement error using SimEx with the rational linear extrapolant was investigated in a simulation study limited to the situation in which the correlation between exposure and each confounder is 0.5, and both confounders are included in the analysis and have an ICC of 0.5. When the confounders are uncorrelated, SimEx correction with the rational linear extrapolant overcorrects for measurement error and produces an estimated exposure-outcome OR of 0.37. Table 5.2 shows that the naïve estimated exposure-outcome OR in this case is 1.56, while the SimEx corrected estimate, using the quadratic extrapolant, is 1.37. The SimEx corrected estimate when using the rational linear extrapolant is more extreme than the corrected estimate using the quadratic extrapolant and the naïve estimate. When the correlation between

the confounders is 0.5, the SimEx corrected estimate using the rational linear extrapolant is 0.89. The SimEx corrected estimate using the quadratic extrapolant in this case is 1.19, while the naïve estimate is 1.39 (see Table 5.3). The overcorrection of measurement error when using the rational linear extrapolant is not as severe when the confounders have a correlation of 0.5 as when the confounders are uncorrelated. In neither of these situations, however, has the SimEx method recovered the true exposure-outcome OR of 1.00.

Figure 5.2 shows the results of using SimEx with the rational linear extrapolant for a single simulated dataset from the analysis in which the correlation between exposure and each confounder is 0.5, and both confounders are included in the analysis and have an ICC of 0.5. For this particular simulated dataset, the naïve exposure-outcome OR is 1.56 and the SimEx corrected OR is 0.41. Considering the estimated exposure-outcome log ORs for $\lambda \geq 0$, increasing measurement error in the confounders has a small effect on the estimated exposure-outcome log OR. This may be the reason that the rational linear extrapolant does not provide an adequate correction for measurement error in this analysis, as the parameters of the extrapolant function may be nearly unidentifiable.[65]

Figure 5.2: Results of simulation extrapolation using the rational linear extrapolant for a single simulated dataset with two confounders in which the correlations between the exposure and the confounders are 0.5, the correlation between the confounders is zero, and both confounders are measured with intra-class correlation coefficient equal to 0.5.



5.3.1.2 Regression calibration

Table 5.4 shows the results of regression calibration correction for simulations with two uncorrelated confounders.

Table 5.4: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, sample size=500,000, repetitions=50. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 — OR adjusted for Z_1	0.1 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.75	0.1 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 1	0.3 — OR adjusted for Z_1	0.3 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.3 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.75	0.3 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 1	0.5 — OR adjusted for Z_1	0.5 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.5 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.75	0.5 — OR adjusted for Z_1 and Z_2, ICC of Z_2 = 1
0.1	0.5	**1.07** *1.10*	**1.00** *1.07*	**1.00** *1.05*	**1.00** *1.04*	**1.22** *1.25*	**1.00** *1.15*	**1.00** *1.10*	**1.00** *1.04*	**1.39**[c] *1.43*[b]	**1.00** *1.26*	**1.00** *1.16*	**1.00** *1.05*
	0.75	**1.07** *1.09*	**1.00** *1.05*	**1.00** *1.04*	**1.00** *1.02*	**1.22** *1.24*	**1.00** *1.13*	**1.00** *1.08*	**1.00** *1.02*	**1.39** *1.42*	**1.00** *1.24*	**1.00** *1.14*	**1.00** *1.02*
	1	- *1.07*	**1.00** *1.04*	**1.00** *1.02*	- *1.00*	- *1.22*	**1.00** *1.11*	**1.00** *1.06*	- *1.00*	- *1.40*[a]	**1.00** *1.22*	**1.00** *1.11*	- *1.00*
0.3	0.5	**1.07** *1.18*	**1.00** *1.15*	**1.00** *1.13*	**1.00** *1.11*	**1.24** *1.36*	**1.00** *1.24*	**1.00** *1.18*	**1.00** *1.12*	**1.43** *1.56*	**1.00** *1.38*	**1.00** *1.27*	**1.00** *1.16*
	0.75	**1.08** *1.13*	**1.00** *1.10*	**1.00** *1.08*	**1.00** *1.06*	**1.24** *1.30*	**1.00** *1.18*	**1.00** *1.13*	**1.00** *1.06*	**1.44** *1.50*	**1.00** *1.31*	**1.00** *1.20*	**1.00** *1.08*
	1	- *1.08*	**1.00** *1.04*	**1.00** *1.02*	- *1.00*	- *1.24*	**1.00** *1.12*	**1.00** *1.06*	- *1.00*	- *1.44*	**1.00** *1.24*	**1.00** *1.13*	- *1.00*
0.5	0.5	**1.09** *1.29*	**1.00** *1.26*	**1.00** *1.24*	**1.00** *1.22*	**1.30** *1.51*	**1.00** *1.38*	**1.00** *1.31*	**1.00** *1.24*	**1.55** *1.76*	**1.00** *1.56*	**1.00** *1.44*	**1.00** *1.31*
	0.75	**1.09** *1.20*	**1.00** *1.16*	**1.00** *1.14*	**1.00** *1.11*	**1.30** *1.41*	**1.00** *1.27*	**1.00** *1.20*	**1.00** *1.13*	**1.56** *1.67*	**1.00** *1.44*	**1.00** *1.31*	**1.00** *1.17*
	1	- *1.09*	**1.00** *1.05*	**1.00** *1.02*	- *1.00*	- *1.30*	**1.00** *1.16*	**1.00** *1.08*	- *1.00*	- *1.56*	**1.00** *1.31*	**1.00** *1.17*	- *1.00*

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0059.*

[a, b, c] *These results are referred to in the text.*

Table 5.5: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratio for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, sample size=500,000, repetitions=50. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1				0.3				0.5			
		OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
			ICC of Z_2				ICC of Z_2				ICC of Z_2		
			0.5	0.75	1		0.5	0.75	1		0.5	0.75	1
0.1	0.5	**1.03**	**1.00**	**1.00**	**1.00**	**1.17**	**1.00**	**1.00**	**1.00**	**1.33**[a]	**1.00**	**1.00**	**1.00**
		1.08	*1.05*	*1.04*	*1.02*	*1.23*	*1.12*	*1.05*	*0.98*	*1.39*[c]	*1.20*	*1.07*	*0.93*
	0.75	**1.03**	**1.00**	**1.00**	**1.00**	**1.18**	**1.00**	**1.00**	**1.00**	**1.34**	**1.00**	**1.00**	**1.00**
		1.06	*1.04*	*1.03*	*1.01*	*1.21*	*1.11*	*1.05*	*0.99*	*1.38*	*1.21*	*1.10*	*0.96*
	1	-	**1.00**	**1.00**	-	-	**1.00**	**1.00**	-	-	**1.00**	**1.00**	-
		1.03	*1.02*	*1.01*	*1.00*	*1.18*	*1.11*	*1.06*	*1.00*	*1.36*[b]	*1.22*	*1.13*	*1.00*
0.3	0.5	**0.97**	**1.00**	**1.00**	**1.00**	**1.11**	**1.00**	**1.00**	**1.00**	**1.28**	**1.00**	**1.00**	**1.00**
		1.12	*1.12*	*1.11*	*1.11*	*1.28*	*1.18*	*1.13*	*1.07*	*1.46*	*1.27*	*1.16*	*1.03*
	0.75	**0.97**	**1.00**	**1.00**	**1.00**	**1.11**	**1.00**	**1.00**	**1.00**	**1.28**	**1.00**	**1.00**	**1.00**
		1.04	*1.05*	*1.05*	*1.06*	*1.20*	*1.13*	*1.09*	*1.04*	*1.38*	*1.23*	*1.13*	*1.02*
	1	-	**1.00**	**1.00**	-	-	**1.00**	**1.00**	-	-	**1.00**	**1.00**	-
		0.97	*0.98*	*0.99*	*1.00*	*1.12*	*1.07*	*1.04*	*1.00*	*1.29*	*1.18*	*1.10*	*1.00*
0.5	0.5	**0.88**	**1.00**	**1.00**	**1.00**	**1.04**	**1.00**	**1.00**	**1.00**	**1.24**	**1.00**	**1.00**	**1.00**
		1.17	*1.20*	*1.21*	*1.22*	*1.36*	*1.27*	*1.23*	*1.18*	*1.57*	*1.39*	*1.27*	*1.15*
	0.75	**0.88**	**1.00**	**1.00**	**1.00**	**1.05**	**1.00**	**1.00**	**1.00**	**1.24**	**1.00**	**1.00**	**1.00**
		1.03	*1.07*	*1.10*	*1.13*	*1.21*	*1.16*	*1.13*	*1.10*	*1.42*	*1.27*	*1.18*	*1.08*
	1	-	**1.00**	**1.00**	-	-	**1.00**	**1.00**	-	-	**1.00**	**1.00**	-
		0.88	*0.93*	*0.96*	*1.00*	*1.05*	*1.03*	*1.02*	*1.00*	*1.25*	*1.15*	*1.08*	*1.00*

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0044.*

[a, b, c] *These results are referred to in the text.*

The regression calibration method removes all bias when both confounders are included in the analysis. Regression calibration occasionally overcorrects for measurement error when only one confounder is included in the analysis. This can be seen when the regression calibration corrected estimate is less than the naïve estimate with a perfectly measured confounder. The overcorrection, however, is small, with a difference between the naïve estimate and the regression calibration estimate of 0.01. The corrected estimates are still biased by unmeasured confounding. This is expected, as the regression calibration method only corrects for measurement errors.

Table 5.5 shows the results of regression calibration correction when the two confounders have a correlation of 0.5. Again, regression calibration correction removes all bias due to residual confounding when both confounders are included in the analysis. When only one confounder is included in the analysis, the regression calibration method tends to overcorrect more frequently when the confounders are correlated than when they are uncorrelated. In Table 5.4, five of the regression calibration corrected estimates when one confounder is included in the analysis are overcorrected for measurement error, compared with ten in Table 5.5. The size of the overcorrection may also be larger when the confounders are correlated than when the confounders are uncorrelated. The largest overcorrection observed in Table 5.4 results in a difference between the regression calibration corrected estimate and the estimate obtained in the absence of measurement error of 0.01. Consider, for example, the situation where the correlation between E and X_1 is 0.1 and the correlation between E and X_2 is 0.5. When only Z_1 is included in the analysis and is measured without error (ICC=1), the estimated exposure-outcome OR is 1.40 (labelled a in Table 5.4). The naïve estimate when Z_1 is measured with ICC equal to 0.5 is 1.43 (labelled b in Table 5.4), which is larger than the estimate obtained without measurement error. The regression calibration corrected estimate when the ICC of Z_1 equals 0.5 is 1.39 (labelled c in Table 5.4), which is smaller than the naïve estimate when Z_1 is measured without error. In comparison, the corresponding regression calibration corrected estimate in Table 5.5 is 1.33 (labelled a in Table 5.5), compared with a naïve estimated OR of 1.36 (labelled b in Table 5.5) when the confounder is measured without error, and a naïve estimated OR of 1.39 (labelled c in Table 5.5) when the ICC of Z_1 equals 0.5. The overcorrection in this situation results in a difference between the regression calibration corrected estimate and the estimate obtained in the absence of measurement error of 0.03.

5.3.2. Four confounders

The results of simulations with four confounders are now presented. For all of the results displayed in this section there are at least two unmeasured confounders, and the remaining two confounders may be either measured with error or omitted from the analysis. Once again, the exposure is assumed to be measured without error. The results in this section are only briefly described as residual confounding is relatively unimportant when compared with the effect of

unmeasured confounding in this situation, and therefore the correction methods will never recover the true exposure-outcome OR of 1.00.

5.3.2.1 Simulation-extrapolation

Table 5.6 and Table 5.7 show the SimEx corrected results for simulations in which the correlation between the confounders is zero or 0.5 respectively. The bold numbers show the SimEx corrected estimates, with the corresponding naïve estimates in italics. Analyses in which SimEx correction was not used, because there was no measurement error in the confounders included in the analyses, are indicated by a dash.

Only very rarely does the SimEx method remove all of the bias due to residual confounding. Generally, the SimEx corrected estimate is intermediate between the naïve estimate and the estimated OR in the absence of residual confounding. When the confounders are uncorrelated, this means the SimEx corrected estimates are closer to the true exposure-outcome OR of 1.00 than then naïve estimates. For correlated confounders, this is not necessarily the case, and the SimEx corrected estimates may be more extreme than the naïve estimate. There are examples of overcorrection of the naïve estimate when the confounders are correlated, where the SimEx corrected estimates are further from the naïve estimate than the estimate obtained in the absence of residual confounding, but these instances are rare.

Correcting for measurement error using SimEx with the rational linear extrapolant was investigated in a simulation study limited to the situation in which the correlation between exposure and X_1, X_2 and X_3 is 0.5, the correlation between exposure and X_4 is 0.3, and Z_1 and Z_2 are included in the analysis and have an ICC of 0.5. When the confounders are uncorrelated, SimEx correction with the rational linear extrapolant overcorrects for measurement error and produces an estimated exposure-outcome OR of 2.44. Table 5.6 shows that the naïve estimated exposure-outcome OR in this case is 3.14, while the SimEx corrected estimate, using the quadratic extrapolant, is 3.06. The SimEx corrected estimate when using the rational linear extrapolant is further from the exposure-outcome OR of 2.88 obtained when the confounders are measured without error than the corrected estimate using the quadratic extrapolant and the naïve estimate. When the correlation between the confounders is 0.5, the SimEx corrected estimate using the rational linear extrapolant is 0.94. The SimEx corrected estimate using the quadratic extrapolant in this case is 1.43, while the naïve estimate is 1.71 (see Table 5.7). The overcorrection of measurement error when using the rational linear extrapolant is not as severe when the confounders have a correlation of 0.5 as when the confounders are uncorrelated, and produces a corrected estimate that is closer to the estimate obtained when the confounders are measured without error than the SimEx corrected estimate using the quadratic extrapolant, or the naïve estimate.

Table 5.6: Geometric means of the simulation-extrapolation corrected estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0, sample size=500,000, repetitions=50. Numbers in bold are the simulation-extrapolation corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1				Correlation between E and X_2 = 0.3				Correlation between E and X_2 = 0.5			
		OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
			ICC of Z_2				ICC of Z_2				ICC of Z_2		
			0.5	0.75	1		0.5	0.75	1		0.5	0.75	1
0.1	0.5	**1.75**	**1.71**	**1.70**	**1.69**	**2.00**	**1.84**	**1.79**	**1.78**	**2.32**	**2.12**	**2.05**	**2.03**
		1.76	*1.73*	*1.72*	*1.71*	*2.01*	*1.91*	*1.85*	*1.79*	*2.33*	*2.20*	*2.13*	*2.05*
	0.75	**1.74**	**1.70**	**1.68**	**1.68**	**1.99**	**1.83**	**1.78**	**1.76**	**2.31**	**2.11**	**2.04**	**2.01**
		1.75	*1.72*	*1.71*	*1.69*	*2.00*	*1.89*	*1.84*	*1.78*	*2.32*	*2.19*	*2.11*	*2.03*
	1	-	**1.69**	**1.68**	-	-	**1.82**	**1.78**	-	-	**2.10**	**2.03**	-
		1.73	*1.71*	*1.69*	*1.68*	*1.99*	*1.88*	*1.82*	*1.76*	*2.31*	*2.17*	*2.09*	*2.00*
0.3	0.5	**1.88**	**1.84**	**1.83**	**1.82**	**2.18**	**2.01**	**1.96**	**1.94**	**2.57**	**2.39**	**2.32**	**2.29**
		1.93	*1.91*	*1.89*	*1.88*	*2.24*	*2.12*	*2.06*	*2.00*	*2.63*	*2.52*	*2.45*	*2.37*
	0.75	**1.84**	**1.79**	**1.78**	**1.78**	**2.14**	**1.96**	**1.90**	**1.88**	**2.52**	**2.32**	**2.24**	**2.21**
		1.88	*1.85*	*1.84*	*1.82*	*2.18*	*2.06*	*2.00*	*1.93*	*2.57*	*2.44*	*2.37*	*2.28*
	1	-	**1.78**	**1.76**	-	-	**1.94**	**1.88**	-	-	**2.30**	**2.22**	-
		1.82	*1.79*	*1.78*	*1.76*	*2.12*	*2.00*	*1.93*	*1.87*	*2.51*	*2.37*	*2.29*	*2.19*
0.5	0.5	**2.17**	**2.12**	**2.11**	**2.10**	**2.60**	**2.39**	**2.32**	**2.30**	**3.18**	**3.06**	**3.00**	**2.97**
		2.23	*2.20*	*2.19*	*2.17*	*2.65*	*2.52*	*2.44*	*2.37*	*3.22*	*3.14*	*3.09*	*3.03*
	0.75	**2.11**	**2.05**	**2.04**	**2.03**	**2.55**	**2.32**	**2.24**	**2.22**	**3.15**	**3.00**	**2.94**	**2.91**
		2.16	*2.13*	*2.11*	*2.09*	*2.59*	*2.45*	*2.37*	*2.29*	*3.18*	*3.09*	*3.03*	*2.96*
	1	-	**2.03**	**2.01**	-	-	**2.29**	**2.21**	-	-	**2.97**	**2.91**	-
		2.09	*2.05*	*2.03*	*2.00*	*2.53*	*2.37*	*2.28*	*2.19*	*3.14*	*3.03*	*2.96*	*2.88*

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, β_1=β_2=β_3=β_4=ln2. Width of 95% simulation intervals<0.010.*

Table 5.7: Geometric means of the simulation-extrapolation corrected estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, sample size=500,000, repetitions=50. Numbers in bold are the simulation-extrapolation corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1				Correlation between E and X_2 = 0.3				Correlation between E and X_2 = 0.5			
		OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
			ICC of Z_2				ICC of Z_2				ICC of Z_2		
			0.5	0.75	1		0.5	0.75	1		0.5	0.75	1
0.1	0.5	**1.53**	**1.53**	**1.53**	**1.53**	**1.72**	**1.50**	**1.42**	**1.40**	**1.94**	**1.52**	**1.35**	**1.29**
		1.54	*1.53*	*1.53*	*1.53*	*1.71*	*1.56*	*1.47*	*1.37*	*1.92*	*1.63*	*1.44*	*1.22*
	0.75	**1.53**	**1.53**	**1.53**	**1.54**	**1.73**	**1.53**	**1.46**	**1.43**	**1.97**	**1.59**	**1.43**	**1.38**
		1.54	*1.53*	*1.53*	*1.53*	*1.72*	*1.59*	*1.50*	*1.40*	*1.94*	*1.69*	*1.52*	*1.30*
	1	-	**1.53**	**1.54**	-	-	**1.55**	**1.48**	-	-	**1.65**	**1.49**	-
		1.53	*1.53*	*1.53*	*1.54*	*1.74*	*1.62*	*1.54*	*1.45*	*1.98*	*1.77*	*1.62*	*1.42*
0.3	0.5	**1.43**	**1.50**	**1.53**	**1.55**	**1.61**	**1.45**	**1.40**	**1.39**	**1.82**	**1.45**	**1.31**	**1.26**
		1.53	*1.56*	*1.59*	*1.62*	*1.70*	*1.58*	*1.51*	*1.42*	*1.92*	*1.64*	*1.47*	*1.27*
	0.75	**1.34**	**1.42**	**1.46**	**1.48**	**1.52**	**1.40**	**1.35**	**1.34**	**1.73**	**1.42**	**1.29**	**1.25**
		1.42	*1.47*	*1.50*	*1.54*	*1.60*	*1.51*	*1.45*	*1.38*	*1.81*	*1.59*	*1.44*	*1.26*
	1	-	**1.40**	**1.43**	-	-	**1.39**	**1.34**	-	-	**1.42**	**1.30**	-
		1.31	*1.37*	*1.40*	*1.45*	*1.49*	*1.42*	*1.38*	*1.33*	*1.70*	*1.52*	*1.40*	*1.25*
0.5	0.5	**1.37**	**1.52**	**1.59**	**1.65**	**1.56**	**1.45**	**1.42**	**1.42**	**1.79**	**1.43**	**1.30**	**1.27**
		1.54	*1.63*	*1.69*	*1.77*	*1.74*	*1.64*	*1.59*	*1.52*	*1.98*	*1.71*	*1.54*	*1.35*
	0.75	**1.18**	**1.35**	**1.43**	**1.49**	**1.36**	**1.31**	**1.29**	**1.30**	**1.58**	**1.30**	**1.20**	**1.17**
		1.34	*1.44*	*1.52*	*1.62*	*1.53*	*1.47*	*1.44*	*1.40*	*1.75*	*1.54*	*1.41*	*1.25*
	1	-	**1.29**	**1.38**	-	-	**1.26**	**1.25**	-	-	**1.27**	**1.17**	-
		1.12	*1.22*	*1.30*	*1.42*	*1.29*	*1.27*	*1.26*	*1.25*	*1.51*	*1.35*	*1.25*	*1.13*

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, β_1=β_2=β_3=β_4=ln2. Width of 95% simulation intervals<0.0048.*

5.3.2.2 Regression calibration

Table 5.8 and Table 5.9 show the regression calibration corrected results for simulations in which the correlation between the confounders is zero or 0.5 respectively. The bold numbers show the regression calibration corrected estimates, with the corresponding naïve estimates in italics. Analyses in which regression calibration correction was not used, because there was no measurement error in the confounders included in the analyses, are indicated by a dash.

In contrast with the results observed for the SimEx corrected estimates, in general the regression calibration method overcorrects the naïve estimate when there are unmeasured confounders. This results in regression calibration corrected estimates that are further from the naïve estimate than the estimate obtained when the confounders are measured without error. When the confounders are uncorrelated, the regression calibration corrected estimates are therefore generally closer to the true exposure-outcome OR of 1.00 than the naïve estimate. Consider, for example, the situation in Table 5.8 when the correlations between the exposure and X_1, and between the exposure and X_2, are 0.1. When both Z_1 and Z_2 are included in the analysis, and both are measured with ICC equal to 0.5, the naïve estimate is 1.73 (labelled a in Table 5.8). The regression calibration corrected estimate is 1.64 (labelled b in Table 5.8), which is closer to the true exposure-outcome OR of 1.00 than the naïve estimate, but further from the naïve estimate than the estimate of 1.68 (labelled c in Table 5.8) obtained when both confounders are measured without error. Regression calibration has overcorrected the naïve estimate, but the result is a corrected estimate that is closer to the true exposure-outcome OR and therefore the overcorrection may not be considered a problem.

It is also possible for the regression calibration corrected estimates to be further from the estimate obtained in the absence of measurement error than the naïve estimate. This occurs, for example, in Table 5.9 when the correlations between each confounder and the exposure are 0.1. When both Z_1 and Z_2 are adjusted for in the analysis, the estimate obtained in the absence of measurement error is 1.54 (labelled a in Table 5.9). When both confounders are measured with error with an ICC of 0.5, the naïve estimate is 1.53 (labelled b in Table 5.9). In this situation, the regression calibration corrected estimate is 1.43 (labelled c in Table 5.9), which is further from the estimate obtained in the absence of measurement error than the naïve estimate. The regression calibration corrected estimate is, however, closer to the true exposure-outcome OR than either the estimate obtained in the absence of measurement error or the naïve estimate. The fact that the regression calibration method has adjusted the estimate away from the estimate obtained in the absence of measurement error may not, therefore, be considered a problem.

The regression calibration method may, however, produce corrected estimates that are more

Table 5.8: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1				Correlation between E and X_2 = 0.3				Correlation between E and X_2 = 0.5			
		OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
			ICC of Z_2				ICC of Z_2				ICC of Z_2		
			0.5	0.75	1		0.5	0.75	1		0.5	0.75	1
0.1	0.5	**1.71** *1.76*	**1.64**[b] *1.73*[a]	**1.65** *1.72*	**1.66** *1.71*	**1.96** *2.01*	**1.72** *1.91*	**1.73** *1.85*	**1.74** *1.79*	**2.26** *2.33*	**1.97** *2.20*	**1.97** *2.13*	**1.97** *2.05*
	0.75	**1.72** *1.75*	**1.65** *1.72*	**1.66** *1.71*	**1.67** *1.69*	**1.97** *2.00*	**1.73** *1.89*	**1.74** *1.84*	**1.75** *1.78*	**2.29** *2.32*	**1.98** *2.19*	**1.98** *2.11*	**1.99** *2.03*
	1	- *1.73*	**1.66** *1.71*	**1.67** *1.69*	- *1.68*[c]	- *1.99*	**1.75** *1.88*	**1.75** *1.82*	- *1.76*	- *2.31*	**2.00** *2.17*	**2.00** *2.09*	- *2.00*
0.3	0.5	**1.81** *1.93*	**1.72** *1.91*	**1.73** *1.89*	**1.75** *1.88*	**2.10** *2.24*	**1.84** *2.12*	**1.84** *2.06*	**1.85** *2.00*	**2.48** *2.63*	**2.17** *2.52*	**2.17** *2.45*	**2.17** *2.37*
	0.75	**1.82** *1.88*	**1.73** *1.85*	**1.74** *1.84*	**1.75** *1.82*	**2.11** *2.18*	**1.84** *2.06*	**1.85** *2.00*	**1.86** *1.93*	**2.49** *2.57*	**2.18** *2.44*	**2.18** *2.37*	**2.18** *2.28*
	1	- *1.82*	**1.74** *1.79*	**1.75** *1.78*	- *1.76*	- *2.12*	**1.85** *2.00*	**1.86** *1.93*	- *1.87*	- *2.51*	**2.19** *2.37*	**2.19** *2.29*	- *2.19*
0.5	0.5	**2.08** *2.23*	**1.96** *2.20*	**1.98** *2.19*	**2.00** *2.17*	**2.53** *2.65*	**2.17** *2.52*	**2.17** *2.44*	**2.18** *2.37*	**3.14** *3.22*	**2.88** *3.14*	**2.88** *3.09*	**2.88** *3.03*
	0.75	**2.08** *2.16*	**1.97** *2.13*	**1.98** *2.11*	**2.00** *2.09*	**2.53** *2.59*	**2.17** *2.45*	**2.18** *2.37*	**2.19** *2.29*	**3.14** *3.18*	**2.88** *3.09*	**2.88** *3.03*	**2.88** *2.96*
	1	- *2.09*	**1.97** *2.05*	**1.99** *2.03*	- *2.00*	- *2.53*	**2.17** *2.37*	**2.18** *2.28*	- *2.19*	- *3.14*	**2.88** *3.03*	**2.88** *2.96*	- *2.88*

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.014.*

[a, b, c] These results are referred to in the text.

Table 5.9: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.5, sample size=500,000, repetitions=50. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1				0.3				0.5			
		OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
			ICC of Z_2				ICC of Z_2				ICC of Z_2		
			0.5	0.75	1		0.5	0.75	1		0.5	0.75	1
0.1	0.5	**1.45** *1.54*	**1.43**[c] *1.53*[b]	**1.46** *1.53*	**1.49** *1.53*	**1.61** *1.71*	**1.37** *1.56*	**1.39** *1.47*	**1.41** *1.37*	**1.81** *1.92*	**1.34** *1.63*	**1.36** *1.44*	**1.38** *1.22*
	0.75	**1.49** *1.54*	**1.46** *1.53*	**1.48** *1.53*	**1.51** *1.53*	**1.67** *1.72*	**1.39** *1.59*	**1.41** *1.50*	**1.43** *1.40*	**1.88** *1.94*	**1.37** *1.69*	**1.38** *1.52*	**1.39** *1.30*
	1	- *1.53*	**1.49** *1.53*	**1.51** *1.53*	- *1.54*[a]	- *1.74*	**1.42** *1.62*	**1.43** *1.54*	- *1.45*	- *1.98*	**1.40** *1.77*	**1.40** *1.62*	- *1.42*
0.3	0.5	**1.27** *1.53*	**1.37** *1.56*	**1.39** *1.59*	**1.42** *1.62*	**1.43** *1.70*	**1.27** *1.58*	**1.29** *1.51*	**1.30** *1.42*	**1.61** *1.92*	**1.21** *1.64*	**1.22** *1.47*	**1.23** *1.27*
	0.75	**1.29** *1.42*	**1.38** *1.47*	**1.41** *1.50*	**1.43** *1.54*	**1.45** *1.60*	**1.29** *1.51*	**1.30** *1.45*	**1.31** *1.38*	**1.65** *1.81*	**1.22** *1.59*	**1.23** *1.44*	**1.24** *1.26*
	1	- *1.31*	**1.41** *1.37*	**1.43** *1.40*	- *1.45*	- *1.49*	**1.30** *1.42*	**1.31** *1.38*	- *1.33*	- *1.70*	**1.23** *1.52*	**1.24** *1.40*	- *1.25*
0.5	0.5	**1.10** *1.54*	**1.34** *1.63*	**1.37** *1.69*	**1.40** *1.77*	**1.26** *1.74*	**1.21** *1.64*	**1.22** *1.59*	**1.23** *1.52*	**1.45** *1.98*	**1.11** *1.71*	**1.12** *1.54*	**1.12** *1.35*
	0.75	**1.11** *1.34*	**1.36** *1.44*	**1.38** *1.52*	**1.40** *1.62*	**1.27** *1.53*	**1.22** *1.47*	**1.23** *1.44*	**1.24** *1.40*	**1.47** *1.75*	**1.12** *1.54*	**1.12** *1.41*	**1.13** *1.25*
	1	- *1.12*	**1.38**[e] *1.22*[d]	**1.39** *1.30*	- *1.42*[f]	- *1.29*	**1.23** *1.27*	**1.24** *1.26*	- *1.25*	- *1.51*	**1.12** *1.35*	**1.13** *1.25*	- *1.13*

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E = 0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0049.*

[a, b, c, d, e, f] These results are referred to in the text.

extreme than the naïve estimate. This occurs, for example, in Table 5.9 when the correlation between exposure and X_1 is 0.5 and the correlation between exposure and X_2 is 0.1. When Z_1 and Z_2 are included in the analysis, and measured with ICCs of 1 and 0.5 respectively, the naïve estimated OR is 1.22 (labelled d in Table 5.9). The regression calibration corrected estimate is 1.38 (labelled e in Table 5.9), which is intermediate between the naïve estimate and the estimate of 1.42 (labelled f in Table 5.9) obtained when Z_1 and Z_2 are measured without error. In this situation, regression calibration has not overcorrected for measurement error, but the corrected estimate is further from the true exposure-outcome OR of 1.00.

From the results presented in this section, overcorrection appears more likely when the effects of residual and unmeasured confounding all act in the same direction. The correlation structure of the datasets simulated for this chapter means that the effect of residual and unmeasured confounding acts to increase the exposure-outcome OR away from the true exposure-outcome OR when the partial correlations between the exposure and confounders are positive. When the confounders are correlated, it is possible for the partial correlations between the exposure and confounders to become negative (as shown in Chapter 3) and the effect of residual confounding then acts to decrease the estimated exposure-outcome OR. In this situation, regression calibration has produced a corrected estimate that is intermediate between the naïve estimate and the one obtained in the absence of measurement error.

Consider the example provided by Armstrong[33] for the effects of measurement error and unmeasured confounding on the estimated exposure-outcome relative risk (RR) (see Section 2.5 and Equation 2.3). To correct for the effects of exposure measurement error, assuming no unmeasured confounding, the regression calibration procedure is equivalent to dividing the naïve exposure effect estimate, $\hat{\beta}_E$, by the ICC of exposure, $R = \frac{\sigma_E^2}{\sigma_{Z_E}^2}$ where σ_E^2 is the variance of the perfectly measured exposure, and $\sigma_{Z_E}^2$ is the variance of the imperfectly measured exposure.[65] The corrected exposure-outcome effect estimate is

Equation 5.2: $$\beta_E^{corr} = \frac{R_{E|1}}{R}\beta_E + \frac{\sigma_{Z_E,X_1}}{R\sigma_{Z_E}^2}\beta_1,$$

where $R_{E|1}$ is the conditional reliability of the exposure given the confounder, and σ_{Z_E,X_1} is the covariance between the error prone exposure measurements and the confounder. The circumstances in which correcting for exposure measurement error will result in overcorrection of the naïve estimate, corrected estimates that are intermediate between the naïve estimate and the estimate obtained in the absence of measurement error, or corrected estimates that are further from the estimate obtained in the absence of measurement error than the naïve estimate can be derived from Equation 2.3 and Equation 5.2. Generally, $R_{E|1} \geq R$, with equality only if both the exposure and the confounder are measured without error. If σ_{E,X_1} denotes the

covariance between the perfectly measured exposure and the confounder, overcorrection, when the corrected estimate is further from the naïve estimate than the estimate obtained in the absence of measurement error, occurs when

Equation 5.3:

$$\beta_E < \frac{\sigma^2_E\sigma_{Z_E,X_1} - \sigma^2_{Z_E}\sigma_{E,X_1}}{(1-R_{E|1})\sigma^2_E\sigma^2_{Z_E}}\beta_1 \text{ and } \beta_E < \frac{R\sigma^2_{Z_E}\sigma_{E,X_1} - \sigma^2_E\sigma_{Z_E,X_1}}{(R_{E|1}-R)\sigma^2_E\sigma^2_{Z_E}}\beta_1, \text{ or}$$

$$\beta_E > \frac{\sigma^2_E\sigma_{Z_E,X_1} - \sigma^2_{Z_E}\sigma_{E,X_1}}{(1-R_{E|1})\sigma^2_E\sigma^2_{Z_E}}\beta_1 \text{ and } \beta_E > \frac{R\sigma^2_{Z_E}\sigma_{E,X_1} - \sigma^2_E\sigma_{Z_E,X_1}}{(R_{E|1}-R)\sigma^2_E\sigma^2_{Z_E}}\beta_1.$$

Corrected estimates that are intermediate between the naïve estimate and the estimate obtained in the absence of measurement error occur when

Equation 5.4:

$$\beta_E > \frac{\sigma_{Z_E,X_1}}{R_{E|1}\sigma^2_{Z_E}}\beta_1 \text{ and } \beta_E > \frac{R\sigma^2_{Z_E}\sigma_{E,X_1} - \sigma^2_E\sigma_{Z_E,X_1}}{(R_{E|1}-R)\sigma^2_E\sigma^2_{Z_E}}\beta_1, \text{ or}$$

$$\beta_E < \frac{\sigma_{Z_E,X_1}}{R_{E|1}\sigma^2_{Z_E}}\beta_1 \text{ and } \beta_E < \frac{R\sigma^2_{Z_E}\sigma_{E,X_1} - \sigma^2_E\sigma_{Z_E,X_1}}{(R_{E|1}-R)\sigma^2_E\sigma^2_{Z_E}}\beta_1.$$

The final option is that the corrected estimate is further from the estimate obtained in the absence of measurement error than the naïve estimate. This will occur when

Equation 5.5:

$$-\frac{\sigma_{Z_E,X_1}}{R_{E|1}\sigma^2_{Z_E}}\beta_1 < \beta_E < \frac{\sigma^2_E\sigma_{Z_E,X_1} - \sigma^2_{Z_E}\sigma_{E,X_1}}{(1-R_{E|1})\sigma^2_E\sigma^2_{Z_E}}\beta_1, \text{ or}$$

$$-\frac{\sigma_{Z_E,X_1}}{R_{E|1}\sigma^2_{Z_E}}\beta_1 > \beta_E > \frac{\sigma^2_E\sigma_{Z_E,X_1} - \sigma^2_{Z_E}\sigma_{E,X_1}}{(1-R_{E|1})\sigma^2_E\sigma^2_{Z_E}}\beta_1.$$

Even in this relatively simple example of the effect of correcting for exposure measurement error in the presence of a single unmeasured confounder, there is a complex relationship that determines whether overcorrection occurs or not. In situations with more confounders the relationship between the naïve estimate, the corrected estimate and the estimate obtained in the absence of measurement error is likely to be even more complex, and predicting when overcorrection will occur will be difficult. Note that the relationships that have been derived above do not exactly apply to parameter estimates from a logistic regression analysis, but if the outcome is rare, the effects of measurement error and correction on the parameter estimates are likely to be close to those described above. The overcorrection of the exposure-outcome OR when using regression calibration observed in this chapter is therefore likely to be due to the correlation structure of the simulated datasets, and different correlation structures may produce corrected estimates that are more likely to be intermediate between the naïve estimate and the estimate obtained in the absence of measurement error. Different correlation structures may also result in a greater frequency of corrected estimates that are more extreme than the naïve estimates.

5.3.3. Exposure measured with error

In this section, corrected exposure effect estimates from simulations in which the exposure was

measured with error, with either an ICC of 0.75 or 0.5, are presented. Only simulated datasets with two confounders are presented, as residual confounding in simulations with four confounder is relatively unimportant when compared with the effects of unmeasured confounding. In addition, only regression calibration corrected estimates are presented, as the previous sections have shown that SimEx in general does not remove all of the bias due to measurement error.

Table 5.10 and Table 5.11 show the regression calibration corrected estimates when the confounders are uncorrelated and the ICC of exposure is 0.75 and 0.5 respectively. Table 5.12 and Table 5.13 show the regression calibration corrected estimates when the correlation between the confounders is 0.5 and the ICC of exposure is 0.75 and 0.5 respectively. The bold numbers show the regression calibration corrected estimates, with the corresponding naïve estimates in italics. These four tables show that, if all confounders are included in the regression analysis, regression calibration corrects perfectly for the effects of measurement error in the exposure and confounders and produces estimates that are equal to the true exposure-outcome OR. When there are unmeasured confounders, the method generally overcorrects for the effects of measurement error and results in regression calibration corrected estimates that are further from the naïve estimate than the estimate that would be obtained in the absence of measurement error. When the effects of measurement errors in the different variables act in different directions, overcorrection can also result in corrected estimates that are further from the estimate that would be obtained in the absence of measurement error than the naïve estimate. This can be seen, for example, in Table 5.11. When the correlation between E and X_1 is 0.3 and the correlation between E and X_2 is 0.1, the estimate obtained in the absence of measurement error when only X_1 is included in the analysis is 1.08 (see Table 5.4). The naïve estimate when the ICC of Z_1 is 0.75 is 1.09 (labelled a in Table 5.11), while the regression calibration corrected estimate is 1.11 (labelled b in Table 5.11). In this situation, overcorrection of the effect of measurement error in the exposure has increased the regression calibration corrected estimate, and has outweighed any effect of correcting for measurement error in Z_1.

In the four tables presented in this section, correcting for measurement error in the exposure has resulted in a corrected crude OR that is more extreme than the naïve OR. This demonstrates the importance of considering the effects of measurement error in, or omission of, all of the relevant variables. When there is residual and/or unmeasured confounding, and the exposure is measured with error, the naïve estimate is not necessarily closer to the null value than the true exposure-outcome OR. Correcting for exposure error alone, may result in corrected estimates that are further from the true estimate than the naïve estimate. Davey Smith and Phillips[146] also made the point that if measurement error correction is used, then it should be used to correct for the effects of measurement errors in the exposure and the confounders, rather than for only errors in the exposure.

Table 5.10: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, ICC of exposure=0.75, sample size=500,000, repetitions=50. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 Crude OR	0.1 OR adjusted for Z_1	0.1 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 … = 0.75	0.1 … = 1	0.3 Crude OR	0.3 OR adjusted for Z_1	0.3 … and Z_2, ICC = 0.5	0.3 … = 0.75	0.3 … = 1	0.5 Crude OR	0.5 OR adjusted for Z_1	0.5 … and Z_2, ICC = 0.5	0.5 … = 0.75	0.5 … = 1
0.1	0.5	**1.16** *1.11*	**1.08** *1.09*	**1.00** *1.06*	**1.00** *1.04*	**1.00** *1.03*	**1.34** *1.24*	**1.25** *1.21*	**1.00** *1.12*	**1.00** *1.08*	**1.00** *1.03*	**1.55** *1.39*	**1.45** *1.36*	**1.00** *1.21*	**1.00** *1.12*	**1.00** *1.04*
	0.75		**1.08** *1.07*	**1.00** *1.04*	**1.00** *1.03*	**1.00** *1.02*		**1.25** *1.20*	**1.00** *1.11*	**1.00** *1.06*	**1.00** *1.02*		**1.46** *1.35*	**1.00** *1.19*	**1.00** *1.11*	**1.00** *1.02*
	1		**1.08** *1.06*	**1.00** *1.03*	**1.00** *1.02*	**1.00** *1.00*		**1.26** *1.19*	**1.00** *1.09*	**1.00** *1.05*	**1.00** *1.00*		**1.47** *1.33*	**1.00** *1.18*	**1.00** *1.09*	**1.00** *1.00*
0.3	0.5	**1.34** *1.24*	**1.08** *1.15*	**1.00** *1.12*	**1.00** *1.11*	**1.00** *1.09*	**1.55** *1.39*	**1.28** *1.29*	**1.00** *1.20*	**1.00** *1.15*	**1.00** *1.10*	**1.80** *1.55*	**1.51** *1.46*	**1.00** *1.30*	**1.00** *1.21*	**1.00** *1.12*
	0.75		**1.09** *1.11*	**1.00** *1.08*	**1.00** *1.06*	**1.00** *1.05*		**1.28** *1.25*	**1.00** *1.15*	**1.00** *1.10*	**1.00** *1.05*		**1.51** *1.41*	**1.00** *1.24*	**1.00** *1.15*	**1.00** *1.06*
	1		**1.09** *1.06*	**1.00** *1.03*	**1.00** *1.02*	**1.00** *1.00*		**1.28** *1.20*	**1.00** *1.10*	**1.00** *1.05*	**1.00** *1.00*		**1.52** *1.36*	**1.00** *1.19*	**1.00** *1.10*	**1.00** *1.00*
0.5	0.5	**1.55** *1.39*	**1.10** *1.24*	**1.00** *1.21*	**1.00** *1.19*	**1.00** *1.18*	**1.80** *1.55*	**1.35** *1.40*	**1.00** *1.30*	**1.00** *1.24*	**1.00** *1.19*	**2.10** *1.75*	**1.65** *1.59*	**1.00** *1.42*	**1.00** *1.33*	**1.00** *1.22*
	0.75		**1.11** *1.16*	**1.00** *1.12*	**1.00** *1.11*	**1.00** *1.09*		**1.35** *1.32*	**1.00** *1.21*	**1.00** *1.15*	**1.00** *1.10*		**1.66** *1.51*	**1.00** *1.33*	**1.00** *1.23*	**1.00** *1.12*
	1		**1.11** *1.07*	**1.00** *1.04*	**1.00** *1.02*	**1.00** *1.00*		**1.35** *1.23*	**1.00** *1.12*	**1.00** *1.06*	**1.00** *1.00*		**1.67** *1.42*	**1.00** *1.22*	**1.00** *1.12*	**1.00** *1.00*

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0059.*

Table 5.11: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0, ICC of exposure=0.5, sample size=500,000, repetitions=50. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.19	**1.10**	**1.00**	**1.00**	**1.00**	1.42	**1.31**	**1.00**	**1.00**	**1.00**	1.70	**1.57**	**1.00**	**1.00**	**1.00**
		1.09	*1.07*	*1.05*	*1.04*	*1.02*	*1.19*	*1.17*	*1.10*	*1.06*	*1.02*	*1.30*	*1.28*	*1.16*	*1.09*	*1.03*
	0.75		**1.10**	**1.00**	**1.00**	**1.00**		**1.32**	**1.00**	**1.00**	**1.00**		**1.58**	**1.00**	**1.00**	**1.00**
			1.06	*1.04*	*1.02*	*1.01*		*1.16*	*1.09*	*1.05*	*1.01*		*1.27*	*1.15*	*1.08*	*1.01*
	1		**1.10**	**1.00**	**1.00**	**1.00**		**1.32**	**1.00**	**1.00**	**1.00**		**1.59**	**1.00**	**1.00**	**1.00**
			1.05	*1.02*	*1.01*	*1.00*		*1.15*	*1.08*	*1.04*	*1.00*		*1.26*	*1.14*	*1.07*	*1.00*
0.3	0.5	1.42	**1.10**	**1.00**	**1.00**	**1.00**	1.70	**1.35**	**1.00**	**1.00**	**1.00**	2.04	**1.65**	**1.00**	**1.00**	**1.00**
		1.19	*1.12*	*1.10*	*1.09*	*1.08*	*1.30*	*1.23*	*1.16*	*1.12*	*1.08*	*1.43*	*1.35*	*1.22*	*1.16*	*1.09*
	0.75		**1.11**[b]	**1.00**	**1.00**	**1.00**		**1.35**	**1.00**	**1.00**	**1.00**		**1.66**	**1.00**	**1.00**	**1.00**
			1.09[a]	*1.06*	*1.05*	*1.04*		*1.19*	*1.12*	*1.08*	*1.04*		*1.31*	*1.18*	*1.11*	*1.04*
	1		**1.11**	**1.00**	**1.00**	**1.00**		**1.36**	**1.00**	**1.00**	**1.00**		**1.67**	**1.00**	**1.00**	**1.00**
			1.05	*1.02*	*1.01*	*1.00*		*1.16*	*1.08*	*1.04*	*1.00*		*1.27*	*1.14*	*1.07*	*1.00*
0.5	0.5	1.70	**1.13**	**1.00**	**1.00**	**1.00**	2.04	**1.44**	**1.00**	**1.00**	**1.00**	2.46	**1.84**	**1.00**	**1.00**	**1.00**
		1.30	*1.18*	*1.16*	*1.15*	*1.14*	*1.43*	*1.30*	*1.22*	*1.18*	*1.14*	*1.57*	*1.44*	*1.30*	*1.23*	*1.16*
	0.75		**1.13**	**1.00**	**1.00**	**1.00**		**1.44**	**1.00**	**1.00**	**1.00**		**1.85**	**1.00**	**1.00**	**1.00**
			1.12	*1.09*	*1.08*	*1.07*		*1.24*	*1.16*	*1.11*	*1.07*		*1.37*	*1.23*	*1.16*	*1.08*
	1		**1.13**	**1.00**	**1.00**	**1.00**		**1.45**	**1.00**	**1.00**	**1.00**		**1.86**	**1.00**	**1.00**	**1.00**
			1.05	*1.03*	*1.01*	*1.00*		*1.17*	*1.09*	*1.04*	*1.00*		*1.30*	*1.16*	*1.08*	*1.00*

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.010.*

[a, b] These results are referred to in the text.

Table 5.12: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, ICC of exposure=0.75, sample size=500,000, repetitions=50. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 Crude OR	0.1 OR adjusted for Z_1	0.1 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 … 0.75	0.1 … 1	0.3 Crude OR	0.3 OR adjusted for Z_1	0.3 … 0.5	0.3 … 0.75	0.3 … 1	0.5 Crude OR	0.5 OR adjusted for Z_1	0.5 … 0.5	0.5 … 0.75	0.5 … 1
0.1	0.5	**1.14**	**1.04**	**1.00**	**1.00**	**1.00**	**1.31**	**1.20**	**1.00**	**1.00**	**1.00**	**1.51**	**1.39**	**1.00**	**1.00**	**1.00**
		1.11	*1.07*	*1.05*	*1.03*	*1.02*	*1.23*	*1.19*	*1.10*	*1.04*	*0.98*	*1.36*	*1.33*	*1.16*	*1.06*	*0.94*
	0.75		**1.04**	**1.00**	**1.00**	**1.00**		**1.20**	**1.00**	**1.00**	**1.00**		**1.40**	**1.00**	**1.00**	**1.00**
			1.05	*1.03*	*1.02*	*1.01*		*1.17*	*1.09*	*1.05*	*0.99*		*1.31*	*1.17*	*1.08*	*0.97*
	1		**1.04**	**1.00**	**1.00**	**1.00**		**1.21**	**1.00**	**1.00**	**1.00**		**1.42**	**1.00**	**1.00**	**1.00**
			1.03	*1.02*	*1.01*	*1.00*		*1.16*	*1.09*	*1.05*	*1.00*		*1.30*	*1.18*	*1.10*	*1.00*
0.3	0.5	**1.31**	**0.96**	**1.00**	**1.00**	**1.00**	**1.51**	**1.12**	**1.00**	**1.00**	**1.00**	**1.74**	**1.32**	**1.00**	**1.00**	**1.00**
		1.23	*1.10*	*1.10*	*1.09*	*1.09*	*1.36*	*1.23*	*1.15*	*1.11*	*1.06*	*1.51*	*1.38*	*1.22*	*1.13*	*1.02*
	0.75		**0.96**	**1.00**	**1.00**	**1.00**		**1.13**	**1.00**	**1.00**	**1.00**		**1.33**	**1.00**	**1.00**	**1.00**
			1.04	*1.04*	*1.05*	*1.05*		*1.17*	*1.11*	*1.07*	*1.03*		*1.31*	*1.18*	*1.10*	*1.01*
	1		**0.96**	**1.00**	**1.00**	**1.00**		**1.13**	**1.00**	**1.00**	**1.00**		**1.34**	**1.00**	**1.00**	**1.00**
			0.97	*0.98*	*0.99*	*1.00*		*1.10*	*1.06*	*1.03*	*1.00*		*1.24*	*1.14*	*1.08*	*1.00*
0.5	0.5	**1.51**	**0.87**	**1.00**	**1.00**	**1.00**	**1.74**	**1.05**	**1.00**	**1.00**	**1.00**	**2.01**	**1.27**	**1.00**	**1.00**	**1.00**
		1.36	*1.14*	*1.16*	*1.17*	*1.18*	*1.51*	*1.29*	*1.22*	*1.18*	*1.14*	*1.69*	*1.45*	*1.30*	*1.21*	*1.11*
	0.75		**0.86**	**1.00**	**1.00**	**1.00**		**1.05**	**1.00**	**1.00**	**1.00**		**1.28**	**1.00**	**1.00**	**1.00**
			1.02	*1.06*	*1.08*	*1.10*		*1.16*	*1.13*	*1.10*	*1.08*		*1.33*	*1.21*	*1.14*	*1.06*
	1		**0.86**	**1.00**	**1.00**	**1.00**		**1.05**	**1.00**	**1.00**	**1.00**		**1.29**	**1.00**	**1.00**	**1.00**
			0.90	*0.94*	*0.97*	*1.00*		*1.03*	*1.02*	*1.01*	*1.00*		*1.19*	*1.11*	*1.06*	*1.00*

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0063.*

Table 5.13: Geometric means of the regression calibration corrected estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.5, ICC of exposure=0.5, sample size=500,000, repetitions=50. Numbers in bold are the regression calibration corrected estimated odds ratios, while those in italics are the naïve estimates.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.18	**1.04**	**1.00**	**1.00**	**1.00**	1.39	**1.25**	**1.00**	**1.00**	**1.00**	1.65	**1.49**	**1.00**	**1.00**	**1.00**
		1.09	*1.06*	*1.04*	*1.03*	*1.01*	*1.18*	*1.15*	*1.08*	*1.03*	*0.99*	*1.28*	*1.26*	*1.12*	*1.04*	*0.95*
	0.75		**1.04**	**1.00**	**1.00**	**1.00**		**1.26**	**1.00**	**1.00**	**1.00**		**1.51**	**1.00**	**1.00**	**1.00**
			1.04	*1.03*	*1.02*	*1.01*		*1.14*	*1.07*	*1.04*	*0.99*		*1.25*	*1.13*	*1.06*	*0.97*
	1		**1.05**	**1.00**	**1.00**	**1.00**		**1.26**	**1.00**	**1.00**	**1.00**		**1.53**	**1.00**	**1.00**	**1.00**
			1.02	*1.01*	*1.01*	*1.00*		*1.12*	*1.07*	*1.04*	*1.00*		*1.24*	*1.14*	*1.08*	*1.00*
0.3	0.5	1.39	**0.95**	**1.00**	**1.00**	**1.00**	1.65	**1.15**	**1.00**	**1.00**	**1.00**	1.96	**1.40**	**1.00**	**1.00**	**1.00**
		1.18	*1.08*	*1.08*	*1.07*	*1.07*	*1.28*	*1.18*	*1.12*	*1.08*	*1.04*	*1.40*	*1.29*	*1.17*	*1.10*	*1.01*
	0.75		**0.95**	**1.00**	**1.00**	**1.00**		**1.16**	**1.00**	**1.00**	**1.00**		**1.42**	**1.00**	**1.00**	**1.00**
			1.03	*1.03*	*1.04*	*1.04*		*1.13*	*1.08*	*1.06*	*1.02*		*1.24*	*1.14*	*1.08*	*1.01*
	1		**0.95**	**1.00**	**1.00**	**1.00**		**1.16**	**1.00**	**1.00**	**1.00**		**1.43**	**1.00**	**1.00**	**1.00**
			0.98	*0.99*	*0.99*	*1.00*		*1.08*	*1.04*	*1.02*	*1.00*		*1.19*	*1.11*	*1.06*	*1.00*
0.5	0.5	1.65	**0.84**	**1.00**	**1.00**	**1.00**	1.96	**1.06**	**1.00**	**1.00**	**1.00**	2.33	**1.34**	**1.00**	**1.00**	**1.00**
		1.28	*1.11*	*1.12*	*1.13*	*1.14*	*1.40*	*1.22*	*1.17*	*1.14*	*1.11*	*1.53*	*1.34*	*1.22*	*1.16*	*1.08*
	0.75		**0.83**	**1.00**	**1.00**	**1.00**		**1.06**	**1.00**	**1.00**	**1.00**		**1.35**	**1.00**	**1.00**	**1.00**
			1.02	*1.04*	*1.06*	*1.08*		*1.12*	*1.10*	*1.08*	*1.06*		*1.24*	*1.16*	*1.10*	*1.04*
	1		**0.83**	**1.00**	**1.00**	**1.00**		**1.07**	**1.00**	**1.00**	**1.00**		**1.37**	**1.00**	**1.00**	**1.00**
			0.92	*0.95*	*0.97*	*1.00*		*1.03*	*1.01*	*1.01*	*1.00*		*1.14*	*1.08*	*1.04*	*1.00*

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.011.*

5.4. Discussion

5.4.1. Summary of results

In this chapter, SimEx and regression calibration were used to correct some of the estimates obtained from the naïve analyses presented in Chapter 3 for measurement error.

Simulation-extrapolation removed some of the residual confounding bias from the naïve estimates. Full correction, however, was rarely observed. Carroll, Ruppert and Stefanski[65] noted that using the quadratic extrapolant tends to produce conservative corrections for the effects of measurement error. This may explain the results seen in this chapter. Cook and Stefanski[81] provided an example showing successful correction for the effects of measurement error in a logistic regression model when using the rational linear extrapolant, with the quadratic and linear extrapolants providing conservative corrections. A limited number of simulations were performed to investigate the effect of using SimEx with the rational linear extrapolant, and this method appeared to overcorrect the estimates, with the overcorrection being worse when the confounders were uncorrelated and producing corrected estimates that were no closer to the estimate obtained in the absence of measurement error than the SimEx corrected estimate using the quadratic extrapolant, or the naïve estimate in which no consideration is given to the effects of measurement error.

The trends observed in the SimEx corrected estimates were driven by the underlying structure of the confounding. This means that the observed trends in the estimated exposure-outcome OR shown in Chapter 3 were still observed for the SimEx corrected estimates. The SimEx method very rarely resulted in an overcorrection of the naïve estimates, where the SimEx corrected estimates were further from the naïve estimate than the estimate obtained when the confounders were measured without error. This may also be due to using the quadratic extrapolant, as overcorrection is less likely when the correction for measurement error is conservative.

Regression calibration was extremely successful in correcting for measurement error of the kind generated in the simulation studies when there was no unmeasured confounding. In contrast with the SimEx corrected estimates, the regression calibration commonly overcorrected for the effects of measurement error when there were unmeasured confounders in the analysis. The expressions derived in Equation 5.3, Equation 5.4 and Equation 5.5 indicate that the circumstances under which overcorrection occurs has a complex relationship with the covariance structure of the data. In the simulations considered in this chapter overcorrection may be considered a minor issue, as it generally results in corrected estimates that are closer to the true exposure-outcome OR.

For both methods, correction may result in a corrected estimate that is further from the true exposure-outcome OR than the naïve estimate when the confounders are correlated. This is due to the effects of negative partial correlations between the exposure and confounders, described in Chapter 3, where residual confounding acts to decrease the overall bias in the estimated exposure-outcome OR. The correction methods will therefore attempt to counteract this effect, and the corrected estimates will be more biased than the naïve estimates, when compared with the true exposure-outcome OR.

5.4.2. Strengths and weaknesses

Only one example of a simulation study comparing regression calibration and SimEx was located in the published literature.[83] The results presented in this chapter extend the results for additive measurement error presented by Fung and Krewski[83] to the situation with multiple explanatory variables measured with error, and to cases in which there are unmeasured confounders. In this chapter, the performances of both SimEx and regression calibration as measurement error correction methods have therefore been investigated in situations closer to those that would be observed in a real-life epidemiological study, where there are likely to be a large number of confounders.

The measurement error generated in these simulation studies had a simple structure. Errors were assumed to be random, non-differential and normally distributed with no correlation between the errors or other variables. The results presented in this chapter are therefore only applicable to this situation, and generalisations should not be made about the performance of either regression calibration or SimEx when the errors do not follow the structure of the simulations.

Cook and Stefanski[81] recommend using a rational linear extrapolant. Due to the instability of this extrapolant function (see Section 4.3.1.2), and the large number of simulations required to produce the results in this chapter, the quadratic extrapolant function was used instead to minimise analysis problems.

5.4.3. Implications

Both SimEx and regression calibration are methods applicable to a range of regression models. They are also easy to use in Stata, with user-written commands[144, 145] available. The results of the simulation studies presented here suggest that regression calibration performs better than SimEx when there are no unmeasured confounders in the setting of random and non-differential measurement error which is uncorrelated with other errors or variables. Regression calibration is also much less computationally intensive than SimEx. These facts imply that regression calibration is a superior method for measurement error correction in exposures and confounders in classical measurement error models, provided there are no unmeasured

confounders. Fung and Krewski[83] also concluded that regression calibration was preferable to SimEx for correcting for measurement error in a single explanatory variable.

5.4.4. Future research

For simulations in which SimEx was used to correct for measurement error, only the quadratic extrapolant was used. The method may have performed better if a different extrapolant, such as a non-linear extrapolant, was used, although the limited simulations performed using the rational linear extrapolant suggested a tendency to overcorrect for measurement error. Further research would be required to investigate this. Only random and non-differential measurement errors were considered. Further work would be required to investigate the relative performance of these two correction methods in more complex measurement error settings.

Only corrections for residual confounding and exposure measurement error when there was no true association between the exposure and outcome was considered in this chapter. As seen in Chapter 3, if the exposure is causally associated with outcome some different trends in the estimated exposure-outcome OR can occur. It would therefore be interesting to evaluate the two correction methods when the exposure is causally related to outcome. These two extensions, however, would not be expected to change the overall trends observed in this chapter. In this setting, SimEx would still be expected to correct some, but not all, of the bias due to residual confounding, with observed trends in the corrected estimates driven by the residual and unmeasured confounding. Regression calibration would be expected to remove all of the bias due to residual confounding when all confounders are included in the analysis, with some overcorrection in situations with unmeasured confounders.

In this chapter, corrections have been based on the known measurement error variances used to simulate the datasets. As these are unlikely to be known in an epidemiological study, a further extension would be to simulate datasets with an internal or external validation study, or a dataset with repeated error-prone measurements, to investigate the performance of regression calibration and SimEx in these more complex, but more realistic, settings.

Chapter 4 showed that there are a wide variety of methods available to correct for measurement error in epidemiological studies. Only two of these methods have been considered in this chapter. Other methods could be investigated in a similar way. These methods would also be expected to perform well in situations in which there is no unmeasured confounding, but not as well when there are unmeasured confounders.

BLANK IN ORIGINAL

Chapter 6.

Background literature: the association between C-reactive protein and coronary heart disease

6.1. Introduction

C-reactive protein (CRP) is a marker of low-grade systemic inflammation. It has recently received attention in the literature, investigating its association with coronary heart disease (CHD) events. One hypothesis is that CRP is a predictor of the stability of atherosclerotic plaques.[147, 148] It remains uncertain whether CRP is causally linked with CHD events, whether it is a marker of sub-clinical disease, or whether the estimated associations are a result of measurement error or unmeasured confounding.

The literature described in this chapter was found by searching Medline for articles with the medical subject headings C-reactive protein and coronary disease or cardiovascular diseases. Only articles in English were included. Titles and abstracts of the articles were reviewed, and those relevant to this literature review were obtained. Studies on populations with specific diseases at baseline, such as diabetes mellitus, were not included. Reference lists of the articles obtained were reviewed, and any further papers relevant to this review were acquired. The literature below is separated by study type, with sections describing estimated associations from cross-sectional studies, nested case-control studies, case-cohort studies, prospective cohort studies and meta-analyses. The chapter concludes with a summary of the published associations.

6.2. The association between CRP and CHD

6.2.1. Cross-sectional studies

Data from the Turkish Adult Risk Factor Study was used by Onat, Sansoy, Yildirim *et al.*[149] Analyses included 744 men and women. Adjustment for age, waist circumference, fibrinogen, total cholesterol and physical activity resulted in an odds ratio (OR) for prevalent CHD of 4.24 (1.63-11.04) when comparing the top quartile of CRP with the bottom. Comparing the third quartile to the bottom gave an estimated OR for prevalent CHD of 2.98 (1.13-7.85), and comparing the second quartile to the bottom quartile gave an estimated OR for prevalent CHD of 1.47 (0.51-4.25). In an analyses adjusted for age, sex, height, systolic blood pressure, fibrinogen and high density lipoprotein (HDL) cholesterol, the OR for prevalent CHD comparing the top quartile to the bottom quartile of CRP was 5.61 (2.04-15.44). Comparing the third quartile of CRP with the bottom quartile gave an OR for prevalent CHD of 4.07 (1.45-11.41), and comparing the second quartile of CRP to the bottom quartile gave an OR for prevalent CHD of 1.80 (0.59-5.47).

Jousilahti, Salomaa, Rasi *et al.*[150] used data from the Finnish Platelet Aggregation and Inflammation Study to assess the association between CRP and prevalent CHD. A total of 1,400

men were analysed. Three models were used to investigate the association between CRP and CHD. In model one, analyses were adjusted for age, smoking, cholesterol, blood pressure and body mass index (BMI). For this analysis, the OR for CHD comparing the top quartile for CRP with the bottom quartile was 2.65 (1.58-4.46). Comparing the third quartile to the bottom quartile of CRP gave an OR for CHD of 2.23 (1.31-3.79), and comparing the second quartile of CRP to the bottom quartile gave an OR for CHD of 1.90 (1.11-3.25). The p-value for trend in this analysis was less than 0.001. Model two adjusted for age, smoking, cholesterol, blood pressure and waist circumference. Comparing the top quartile of CRP with the bottom quartile gave an OR for CHD of 2.33 (1.36-3.99). Comparing the third quartile of CRP with the bottom quartile gave an OR for CHD of 2.05 (1.20-3.50), while comparing the second quartile of CRP with the bottom quartile gave an OR for CHD of 1.82 (1.06-3.12). The p-value for trend in this analysis was 0.003. The third model adjusted for age, smoking, cholesterol, blood pressure and waist hip ratio. Comparing the top quartile of CRP with the bottom quartile gave an OR for CHD of 2.20 (1.29-3.75). Comparing the third quartile of CRP with the bottom quartile gave an OR for CHD of 1.98 (1.16-3.38), while comparing the second quartile of CRP with the bottom quartile gave an OR for CHD of 1.80 (1.05-3.08). The p-value for trend for this analysis was 0.006.

Data from the Monitoring Trends and Determinants in Cardiovascular Disease (MONICA) project in Glasgow, Scotland, were used by Woodward, Rumley, Lowe *et al.*[151] to investigate the cross-sectional association between CRP and cardiovascular disease (CVD). A total of 414 men and 550 women were analysed. Analyses on men were adjusted for age, total cholesterol, triglycerides, diastolic blood pressure, plasma vitamin C, smoking, cotinine and fibrinogen. Analyses on women additionally adjusted for menopause and hormone replacement therapy (HRT) status, and contraceptive pill use. The OR for prevalent CVD, comparing the top quartile of CRP with the bottom, was 1.16 (0.53-2.54) for men and 1.87 (0.90-3.88) for women. Comparing the third quartile to the bottom quartile gave and OR for CHD of 1.02 (0.47-2.22) for men and 1.27 (0.64-2.52) for women, while comparing the second quartile of CRP to the bottom quartile gave an OR for CHD of 0.76 (0.33-1.66) for men and 1.12 (0.56-2.27) for women. The p-values for trend were 0.44 for men and 0.09 for women.

Anand, Razak, Yi *et al.*[152] investigated the association between CRP and CVD in a sample of people with South Asian, Chinese, European or Aboriginal ancestry living in Canada. The study included 1,250 people. Logistic regression models were adjusted for age, sex, systolic blood pressure, low density lipoprotein (LDL) cholesterol, HDL cholesterol, smoking, diabetes, atherosclerosis, waist circumference, BMI, triglycerides and ethnicity. The OR for prevalent CHD was 1.03 (1.00-1.06) for a 0.1 mg/L increase in CRP.

Data from the British Women's Heart and Health Study (BWHHS) was used by Lawlor, Davey Smith, Rumley *et al.*[153] in a cross-sectional analysis. Using data from 2,987 women the

Table 6.1: Summary of estimated associations between C-reactive protein and coronary heart disease or cardiovascular disease from cross-sectional studies.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Onat *et al.*[149]	744 subjects	Prevalent CHD	0.9-1.85 mg/L vs. < 0.9 mg/L	ABCDG	1.47	(0.51-4.25)	
			1.86-4.2 mg/L vs. < 0.9 mg/L	ABCDG	2.98	(1.13-7.85)	
			> 4.2 mg/L vs. < 0.9 mg/L	ABCDG	4.24	(1.63-11.04)	
		Prevalent CHD	0.9-1.85 mg/L vs. < 0.9 mg/L	ACDG	1.80	(0.59-5.47)	
			1.86-4.2 mg/L vs. < 0.9 mg/L	ACDG	4.07	(1.45-11.41)	
			> 4.2 mg/L vs. < 0.9 mg/L	ACDG	5.61	(2.04-15.44)	
Jousilahti *et al.*[150]	1,400 men	Prevalent CHD	Second quartile vs. bottom quartile	ABC	1.90	(1.11-3.25)	<0.001
			Third quartile vs. bottom quartile	ABC	2.23	(1.31-3.79)	
			Top quartile vs. bottom quartile	ABC	2.65	(1.58-4.46)	
		Prevalent CHD	Second quartile vs. bottom quartile	ABCD	1.82	(1.06-3.12)	0.003
			Third quartile vs. bottom quartile	ABCD	2.05	(1.20-3.50)	
			Top quartile vs. bottom quartile	ABCD	2.33	(1.36-3.99)	
		Prevalent CHD	Second quartile vs. bottom quartile	ABCD	1.80	(1.05-3.08)	0.006
			Third quartile vs. bottom quartile	ABCD	1.98	(1.16-3.38)	
			Top quartile vs. bottom quartile	ABCD	2.20	(1.29-3.75)	
Woodward *et al.*[151]	414 men	Prevalent CVD	Second quartile vs. bottom quartile	ABCG	0.76	(0.33-1.66)	0.44
			Third quartile vs. bottom quartile	ABCG	1.02	(0.47-2.22)	
			Top quartile vs. bottom quartile	ABCG	1.16	(0.53-2.54)	
	550 women	Prevalent CVD	Second quartile vs. bottom quartile	ABCFG	1.12	(0.56-2.27)	0.09
			Third quartile vs. bottom quartile	ABCFG	1.27	(0.64-2.52)	
			Top quartile vs. bottom quartile	ABCFG	1.87	(0.90-3.88)	
Anand *et al.*[152]	1,250 subjects	Prevalent CVD	0.1 mg/L increase in CRP	ABCD	1.03	(1.00-1.06)	

Table 6.1 continued.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Lawlor *et al.*[153]	2,987 women	Prevalent CHD	Doubling of CRP	ABCDE	1.06	(0.98-1.15)	

Abbreviations: CI, confidence interval; CHD, coronary heart disease; CVD, cardiovascular disease; CRP, C-reactive protein; FEV_1, forced expiratory volume in one second.
A: Adjusted for general factors, such as age and sex.
B: Adjusted for behavioural risk factors, such as smoking and alcohol intake.
C: Adjusted for physiological risk factors, such as blood pressure and cholesterol.
D: Adjusted for anthropometric and/or lung function variables, such as height and FEV_1.
E: Adjusted for socio-economic status.
F: Adjusted for medications, such as statins or antihypertensive drugs.
G: Adjusted for other inflammatory markers, such as fibrinogen or interleukin-6.

association between CRP and prevalent CHD was investigated. Analyses were adjusted for age, life course socio-economic position score, behavioural and physiological risk factors, adult anthropometry and forced expiratory volume in one second (FEV_1). The OR for CHD for a doubling of CRP was 1.06 (0.98-1.15).

The estimated associations described in this section are summarised in Table 6.1. Although many of the studies found a positive association between CRP and CHD or CVD, reverse causation, where CHD or CVD causes elevated levels of CRP, is a possible explanation for these results.

6.2.2. Nested case-control studies

A nested case-control study identifies cases and controls from the participants of an already existing cohort study or randomised trial. One advantage of this type of case-control study over a non-nested case control study is that exposures will often be measured at the beginning of the cohort study or randomised trial within which the study is nested, and therefore before the onset of disease. This will reduce the potential for recall bias, which may be a major problem in non-nested case-control studies.

The Multiple Risk Factor Intervention Trial (MRFIT) was a randomised controlled trial in which a multifactorial intervention was used to prevent CHD morbidity and mortality in men at high risk of CHD but with no clinical evidence of the disease. Kuller, Tracy, Shaten *et al.*[154] used a nested case-control design to investigate the effect of CRP on CHD. The 246 cases comprised 98 participants who had experienced a non-fatal MI, and 148 participants who died due to CHD. The 491 controls were matched with cases by age, smoking status, clinic, and study group (intervention or usual care). All analyses were adjusted for age, number of cigarettes smoked per day, diastolic blood pressure, triglycerides, LDL cholesterol and HDL cholesterol. The OR for CHD death was 2.8 (1.4-5.4) when comparing the top quartile of CRP with the bottom quartile. Comparing the third quartile of CRP to the bottom quartile gave an OR for CHD death of 2.7 (1.4-5.2), and comparing the second quartile of CRP to the bottom quartile gave an OR for CHD death of 1.6 (0.8-3.1). There was little evidence of an association between quartiles of CRP and non-fatal MI. Comparing the second quartile of CRP with the bottom quartile gave an OR for non-fatal MI of 0.6 (0.3-1.3), comparing the third quartile of CRP with the bottom quartile gave an OR for non-fatal MI of 0.9 (0.4-1.9), and comparing the top quartile of CRP with the bottom quartile gave an OR for non-fatal MI of 0.8 (0.4-1.7). When the analysis was restricted to smokers, the OR for CHD death was 4.3 (1.7-10.8) when comparing the top quartile of CRP with the bottom quartile, 3.2 (1.3-7.7) when comparing the third quartile of CRP with the bottom quartile, and 1.7 (0.7-4.3) when comparing the second quartile of CRP with the bottom quartile. There was little evidence of an association between quartiles of CRP and non-fatal MI among the smokers. The OR for non-fatal MI was 1.0 (0.4-2.5) comparing the top quartile of CRP with

the bottom quartile, 1.2 (0.4-3.2) comparing the third quartile of CRP to the bottom quartile, and 0.7 (0.3-1.9) comparing the second quartile of CRP to the bottom quartile.

The Cardiovascular Health Study (CHS) was a population based cohort study of CHD and stroke in adults aged 65 and over. Tracy, Lemaitre, Psaty *et al.*[155] used a prospective nested case-control study of participants of the CHS to investigate the effect of CRP on incident CHD events. A total of 146 cases were matched for sex, presence or absence of sub-clinical disease, and duration of follow-up with 146 controls. Statistical analyses accounted for all matching factors. The OR for incident CHD was 1.07 (0.52-2.22) for men when comparing the top quartile of CRP with the lower three quartiles. For women, the OR for incident CHD was 1.60 (0.73-3.53) when comparing the top quartile of CRP with the lower three quartiles. When analysing only participants with subclinical CVD at baseline, the OR for incident CHD among men was 1.09 (0.49-2.37) when comparing the top quartile of CRP with the bottom quartile, and the OR for incident CHD among women was 2.33 (0.90-6.07) when comparing the top quartile of CRP with the lower three quartiles.

Tracy, Lemaitre, Psaty *et al.*[155] also considered a prospective nested case-control study to assess the effect of CRP on incident cardiovascular events in the Rural Health Promotion Project. The analysis included 145 cases and 146 controls matched for age and sex. For women, the OR for incident CHD was 2.7 (1.10-6.69) when comparing the top quintile of CRP with the lower four quintiles, and the OR for CHD death only was 3.74 (1.36-10.4) comparing the top quintile of CRP with the lower four quintiles. For men, the OR for incident CHD was 2.0 (0.82-4.87) when comparing the top quintile of CRP with the lower four quintiles, and the OR for CHD deaths only was 1.4 (0.50-3.95) when comparing the top quintile of CRP with the lower four quintiles.

The Physicians' Health Study (PHS) was a 2 by 2 factorial randomised controlled trial to assess the effect of aspirin and beta-carotene on prevention of CVD and cancer. Ridker, Cushman, Stampfer *et al.*[156] used a prospective nested case-control study of participants in the PHS to investigate the effect of plasma levels of CRP on incidence of MI, stroke or venous thrombosis. A total of 543 cases were matched for age, smoking status, and length of time since randomisation with 543 controls who had also provided blood samples at baseline. The OR for MI was 2.6 (1.6-4.4) when comparing the top quartile of CRP concentration with the bottom quartile, 2.4 (1.5-4.0) when comparing the third quartile of CRP with the bottom quartile, and 1.5 (0.9-2.5) when comparing the second quartile of CRP with the bottom quartile. The p-value for trend was less than 0.001. This analysis adjusted for BMI, diabetes, history of hypertension and family history of coronary artery disease. No results from analyses in which adjustment for confounders was made were presented for the outcomes of stroke or venous thrombosis.

Ridker, Buring, Shih *et al.*[157] analysed 122 cases and 244 age and smoking matched controls from

the Women's Health Study (WHS). The WHS included post-menopausal female health professionals, and was a RCT designed to investigate the effect of aspirin and vitamin E on CVD and cancer. In models allowing for the matching variables and adjusted for BMI, diabetes, hypertension, exercise, family history of coronary artery disease and treatment assignment, the OR for any cardiovascular event comparing the top quartile with the bottom quartile of CRP was 4.1 (1.7-9.9). Comparing the third quartile of CRP to the bottom quartile, the OR for any cardiovascular event was 2.3 (1.0-5.6), and the OR for any cardiovascular event was 2.0 (0.8-4.7) when comparing the second quartile of CRP to the bottom quartile. The p-value for trend for this analysis was 0.001. When the outcome events were restricted to MI or stroke, the OR for MI or stroke was 5.5 (1.8-16.6) when comparing the top quartile of CRP to the bottom quartile. The OR for MI or stroke comparing the third quartile of CRP to the bottom quartile was 3.5 (1.1-10.4), and comparing the second quartile of CRP to the bottom quartile gave an OR for MI or stroke of 2.7 (0.9-8.1). The p-value for trend for this analysis was 0.002.

Data from the WHS was again used in a nested case-control analysis by Ridker, Hennekens, Buring *et al.*[158] to investigate the association between CRP and cardiovascular events. A total of 122 cases and 244 age and smoking matched controls were analysed. Analyses accounted for the matching variables and were adjusted for random assignment to aspirin or vitamin E, BMI, history of hypertension, history of diabetes, parental history of MI and other plasma markers of inflammation. For a single quartile increase in CRP, the OR for a cardiovascular event was 1.5 (1.1-2.1).

Packard, O'Reilly, Caslake *et al.*[159] conducted a nested case-control study within the West of Scotland Coronary Prevention Study (WOSCOPS). This was a randomised study of 6,595 men with LDL cholesterol between 174 and 232 mg/dL at baseline, but with no history of MI, to investigate the effect of pravastatin on CHD. The nested case-control study included 580 CHD cases, and 1,160 controls matched for age and smoking. Analyses adjusted for age, systolic blood pressure, triglycerides, LDL and HDL cholesterol, fibrinogen, white-cell count and lipoprotein-associated phospholipase A_2. For a standard deviation increase in log CRP, the OR for CHD was 1.13 (0.98-1.29).

The British Regional Heart Study was a prospective cohort study in which 7,735 men from 24 British towns were followed up for all-cause mortality and cardiovascular morbidity. Danesh, Whincup, Walker *et al.*[160] used a nested case-control study design to investigate the effect of CRP, serum amyloid A protein, leukocyte count, and albumin on cardiovascular death and non-fatal MI. A total of 506 cases were matched with 1,025 controls based on town of residence and age in five year bands. Multivariable analyses controlled for age, town, smoking, blood pressure, total cholesterol, HDL cholesterol, BMI, occupation, housing tenure, marital status, car ownership, and childhood socioeconomic factors. The OR for CHD was 2.13 (1.38-3.28)

comparing the top tertile of CRP with the bottom tertile. When this analysis was restricted to men without evidence of CHD at baseline, the OR for CHD was 2.31 (1.42-3.76) when comparing the top tertile of CRP to the bottom tertile. The ORs for CHD comparing the middle tertile of CRP to the bottom tertile were not presented for either analysis.

Roivainen, Viik-Kajander, Palosuo *et al.*[161] used data from the Helsinki Heart Study. This was a placebo-controlled RCT of gemfibrozil in dyslipidemic men on the primary prevention of CHD. The nested case-control study consisted of 241 cases who had suffered either MI or CHD death, matched with 241 controls for treatment group and area of residence. Adjusting for age and smoking, the OR for CHD comparing the top quartile of CRP with the bottom quartile was 3.66 (1.97-6.81). Comparing the third quartile of CRP to the bottom quartile gave an OR for CHD of 1.57 (0.87-2.82), and comparing the second quartile of CRP to the bottom quartile gave an OR for CHD of 0.98 (0.55-1.75).

Pradhan, Manson, Rossouw *et al.*[162] conducted a nested case-control study of post-menopausal women in the observational study component of the Women's Health Initiative. This is a prospective cohort study of post-menopausal women, and is designed to investigate the association between many risk factors (clinical, socioeconomic, behavioural and dietary) and the incidence of several health outcomes, including MI. The nested case-control study consisted of 304 cases who had experienced a first MI during follow-up. The 304 controls were matched with the cases on age, smoking status, ethnicity and follow-up time. The analyses accounted for all matching variables, and additionally adjusted for the ratio of total cholesterol to HDL cholesterol, BMI, history of hypertension, family history of premature coronary artery disease, diabetes, exercise frequency, alcohol consumption, and use of HRT. The OR for incident MI, comparing the top quartile of CRP to the bottom quartile, was 2.1 (1.1-4.1). Comparing the third quartile of CRP to the bottom quartile gave an OR for incident MI of 1.4 (0.7-2.6), and comparing the second quartile of CRP to the bottom quartile gave an OR for incident MI of 1.4 (0.8-2.8). The p-value for trend was 0.046.

The Rotterdam Study is a prospective cohort study of people 55 years and older, designed to investigate risk factors for, and incidence of, chronic disabling diseases. Van der Meer, de Maat, Kiliaan *et al.*[163] conducted a nested case-control study within the Rotterdam Study. The cases were 157 people who experienced an MI during follow-up. The 500 controls were randomly selected from the people who had not had an MI during follow-up, and who had not died. Analyses adjusted for age, age squared, sex, current smoking, BMI, hypertension, diabetes mellitus, family history of early MI, total cholesterol and HDL cholesterol. The OR for incident MI, comparing the top quartile of CRP with the bottom quartile, was 1.2 (0.6-2.2). When comparing the third quartile of CRP to the bottom quartile, the OR for incident MI was 1.0 (0.5-1.9), and when comparing the second quartile of CRP to the bottom quartile, the OR for incident

MI was 0.9 (0.5-1.7). The p-value for trend was 0.50.

The Prospective Epidemiological Study of Myocardial Infarction (PRIME) study is a cohort study in France and Northern Ireland to investigate the association between various risk factors and CHD. Luc, Bard, Juhan-Vague *et al.*[164] used a nested case-control study of PRIME participants to investigate the effect of CRP, interleukin-6 and fibrinogen on CHD. A total of 317 cases were matched for recruitment in the same centre and on the same day with 609 controls. Controls were free of CHD on the date of the event for the case. In a multivariable model in which age, diabetes, smoking, hypertension, LDL cholesterol, HDL cholesterol, and triglycerides were controlled, the OR for incident CHD was 2.16 (1.26-3.72) when comparing the top tertile of CRP with the bottom. Comparing the middle tertile of CRP to the bottom tertile, the OR for incident CHD was 0.81 (0.47-1.40). The p-value for trend was 0.002. When interleukin-6 and fibrinogen were also included in a logistic regression analysis, with adjustment for the same confounders, the OR for incident CHD was 1.02 (0.98-1.06), but the increase in CRP to which this relates is not clear.

Data from the Reykjavik Study, a prospective study of CVD, was used by Danesh, Wheeler, Hirschfield *et al.*[165] in a case-control analysis. A total of 2,459 cases with non-fatal MI or fatal CHD were matched by calendar year of recruitment, sex, and age with 3,969 controls. Analyses were adjusted for age, sex, calendar year of enrolment, smoking, systolic blood pressure, total cholesterol level, triglyceride level, BMI, FEV_1, diabetes, socio-economic status, erythrocyte sedimentation rate and von Willebrand factor level. Comparing the top tertile of CRP with the bottom tertile, the adjusted OR for CHD was 1.36 (1.16-1.58). Restricting this analysis to people without evidence of CHD at baseline, the OR for CHD was 1.30 (1.10-1.52) when comparing the top tertile of CRP to the bottom tertile. The ORs for CHD comparing the middle tertile of CRP to the bottom tertile were not presented for either of these analyses.

Pai, Pischon, Ma *et al.*[166] used a nested case-control design to investigate the association between CRP and CHD in the Nurses' Health Study (NHS), a prospective study of female registered nurses in the U.S., and the Health Professionals Follow-up Study (HPFS), a prospective cohort study of male health professionals. A total of 239 cases and 469 controls, matched for age, smoking status, date of blood sampling and fasting status at the time of blood sampling, were identified from the NHS. From the HPFS, 265 cases and 529 controls, matched for age, smoking status and date of blood sampling, were identified. Analyses investigating the association between quintiles of CRP and CHD were adjusted for the matching factors, parental history of CHD before age 60, alcohol intake, physical activity, ratio of total cholesterol to HDL cholesterol, BMI, diabetes, hypertension and, for the analyses on women, HRT. Among the women, the OR for CHD was 1.61 (0.84-3.07) comparing the top quintile of CHD to the bottom quintile. Comparing the fourth quintile to the bottom quintile, the OR for CHD was 1.22 (0.65-

Table 6.2: Summary of estimated associations between C-reactive protein and coronary heart disease or cardiovascular disease from nested case-control studies.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Kuller *et al.*[154]	246 cases 491 controls	Fatal CHD	Second quartile vs. bottom quartile	ABC	1.6	(0.8-3.1)	
			Third quartile vs. bottom quartile	ABC	2.7	(1.4-5.2)	
			Top quartile vs. bottom quartile	ABC	2.8	(1.4-5.4)	
		Non-fatal MI	Second quartile vs. bottom quartile	ABC	0.6	(0.3-1.3)	
			Third quartile vs. bottom quartile	ABC	0.9	(0.4-1.9)	
			Top quartile vs. bottom quartile	ABC	0.8	(0.4-1.7)	
	100 cases 200 controls (smokers)	Fatal CHD	Second quartile vs. bottom quartile	ABC	1.7	(0.7-4.3)	
			Third quartile vs. bottom quartile	ABC	3.2	(1.3-7.7)	
			Top quartile vs. bottom quartile	ABC	4.3	(1.7-10.8)	
		Non-fatal MI	Second quartile vs. bottom quartile	ABC	0.7	(0.3-1.9)	
			Third quartile vs. bottom quartile	ABC	1.2	(0.4-3.2)	
			Top quartile vs. bottom quartile	ABC	1.0	(0.4-2.5)	
Tracy *et al.*[155]	89 male cases 89 male controls	Incident CHD	2.79-46.04 mg/L vs. ≤ 2.78 mg/L	AC	1.07	(0.52-2.22)	
	57 female cases 57 female controls	Incident CHD	2.79-46.04 mg/L vs. ≤ 2.78 mg/L	AC	1.60	(0.73-3.53)	
	66 male cases 66 male controls (with subclinical CHD)	Incident CHD	2.79-46.04 mg/L vs. ≤ 2.78 mg/L	AC	1.09	(0.49-2.37)	
	41 female cases 41 female controls (with subclinical CHD)	Incident CHD	2.79-46.04 mg/L vs. ≤ 2.78 mg/L	AC	2.33	(0.90-6.07)	

Table 6.2 continued.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Tracy *et al.*[155]	65 female cases 65 or 66 female controls	Incident CHD Fatal CHD	Top quintile vs. lower four quintiles Top quintile vs. lower four quintiles	A A	2.7 3.74	(1.10-6.69) (1.36-10.4)	
	80 male cases 80 or 81 male controls	Incident CHD Fatal CHD	Top quintile vs. lower four quintiles Top quintile vs. lower four quintiles	A A	2.0 1.4	(0.82-4.87) (0.50-3.95)	
Ridker *et al.*[156]	543 cases 543 controls	MI	Second quartile vs. bottom quartile Third quartile vs. bottom quartile Top quartile vs. bottom quartile	ABC ABC ABC	1.5 2.4 2.6	(0.9-2.5) (1.5-4.0) (1.6-4.4)	<0.001
Ridker *et al.*[157]	122 cases 244 controls	Incident CVD	1.5-3.7 mg/L vs. < 1.5 mg/L 3.8-7.3 mg/L vs. < 1.5 mg/L > 7.3 mg/L vs. < 1.5 mg/L	ABC ABC ABC	2.0 2.3 4.1	(0.8-4.7) (1.0-5.6) (1.7-9.9)	0.001
		MI or stroke	1.5-3.7 mg/L vs. < 1.5 mg/L 3.8-7.3 mg/L vs. < 1.5 mg/L > 7.3 mg/L vs. < 1.5 mg/L	ABC ABC ABC	2.7 3.5 5.5	(0.9-8.1) (1.1-10.4) (1.8-16.6)	0.002
Ridker *et al.*[158]	122 cases 244 controls	Incident CVD	Single quartile increase in CRP	ABCG	1.5	(1.1-2.1)	
Packard *et al.*[159]	580 cases 1,160 controls	Incident CHD	Standard deviation increase in log CRP	ACG	1.13	(0.98-1.29)	
Danesh *et al.*[160]	506 cases 1,025 controls	Incident CHD	> 2.4 mg/L vs. < 0.9 mg/L	ABCE	2.13	(1.38-3.28)	
	329 cases 820 controls (without CHD at baseline)	Incident CHD	> 2.4 mg/L vs. < 0.9 mg/L	ABCE	2.31	(1.42-3.76)	

Table 6.2 continued.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Roivainen *et al.*[161]	241 cases 241 controls	Incident CHD	Second quartile vs. bottom quartile	AB	0.98	(0.55-1.75)	
			Third quartile vs. bottom quartile	AB	1.57	(0.87-2.82)	
			Top quartile vs. bottom quartile	AB	3.66	(1.97-6.81)	
Pradhan *et al.*[162]	304 cases 304 controls	Incident MI	>0.10-0.24 mg/dL vs. ≤ 0.10 mg/dL	ABCF	1.4	(0.8-2.8)	0.046
			0.25-0.47 mg/dL vs. ≤ 0.10 mg/dL	ABCF	1.4	(0.7-2.6)	
			> 0.47 mg/dL vs. ≤ 0.10 mg/dL	ABCF	2.1	(1.1-4.1)	
van der Meer *et al.*[163]	157 cases 500 controls	Incident MI	0.82-1.68 mg/L vs. < 0.82 mg/L	ABC	0.9	(0.5-1.7)	0.50
			1.68-3.02 mg/L vs. < 0.82 mg/L	ABC	1.0	(0.5-1.9)	
			> 3.02 mg/L vs. < 0.82 mg/L	ABC	1.2	(0.6-2.2)	
Luc *et al.*[164]	317 cases 609 controls	Incident CHD	0.75- <1.97 mg/L vs.< 0.75 mg/L	ABC	0.81	(0.47-1.40)	0.002
			≥ 1.97 mg/L vs. < 0.75 mg/L	ABC	2.16	(1.26-3.72)	
Danesh *et al.*[165]	2,459 cases 3,969 controls	Incident CHD	> 2.0 mg/L vs. < 0.78 mg/L	ABCDEG	1.36	(1.16-1.58)	
	2,042 cases 3,891 controls (without CHD at baseline)	Incident CHD	> 2.0 mg/L vs. < 0.78 mg/L	ABCDEG	1.30	(1.10-1.52)	
Pai *et al.*[166]	239 female cases 469 female controls	Incident CHD	0.80-1.70 mg/L vs. < 0.80 mg/L	ABCF	1.23	(0.66-2.32)	0.08
			1.71-2.91 mg/L vs. < 0.80 mg/L	ABCF	0.89	(0.46-1.72)	
			2.92-5.96 mg/L vs. < 0.80 mg/L	ABCF	1.22	(0.65-2.30)	
			≥ 5.97 mg/L vs. < 0.80 mg/L	ABCF	1.61	(0.84-3.07)	

Table 6.2 continued.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Pai *et al.*[166]	239 female cases 469 female controls	Incident CHD	1.0-2.9 mg/L vs. < 1.0 mg/L	ABCF	1.17	(0.69-2.00)	0.09
			≥ 3.0 mg/L vs. < 1.0 mg/L	ABCF	1.53	(0.89-2.62)	
		Incident CHD	0.80-1.70 mg/L vs. < 0.80 mg/L	ABC	1.75	(0.97-3.16)	0.02
			1.71-2.91 mg/L vs. < 0.80 mg/L	ABC	1.74	(0.96-3.15)	
			2.92-5.96 mg/L vs. < 0.80 mg/L	ABC	2.14	(1.18-3.88)	
			≥ 5.97 mg/L vs. < 0.80 mg/L	ABC	2.55	(1.40-4.65)	
	265 male cases 529 male controls	Incident CHD	1.0-2.9 mg/L vs. < 1.0 mg/L	ABC	1.60	(1.09-2.34)	0.03
			≥ 3.0 mg/L vs. < 1.0 mg/L	ABC	1.79	(1.14-2.83)	
	504 cases 998 controls	Incident CHD	1.0-2.9 mg/L vs. < 1.0 mg/L	ABC	1.44	(1.05-1.96)	0.008
			≥ 3.0 mg/L vs. < 1.0 mg/L	ABC	1.68	(1.18-2.38)	

Abbreviations: CI, confidence interval; CHD, coronary heart disease; CVD, cardiovascular disease; MI, myocardial infarction; CRP, C-reactive protein; FEV_1, forced expiratory volume in one second.

A: Adjusted for general factors, such as age and sex.
B: Adjusted for behavioural risk factors, such as smoking and alcohol intake.
C: Adjusted for physiological risk factors, such as blood pressure and cholesterol.
D: Adjusted for anthropometric and/or lung function variables, such as height and FEV_1.
E: Adjusted for socio-economic status.
F: Adjusted for medications, such as statins or antihypertensive drugs.
G: Adjusted for other inflammatory markers, such as fibrinogen or interleukin-6.

2.30), and the OR for CHD was 0.89 (0.46-1.72) when comparing the third quintile of CRP to the bottom quintile. Comparing the second quintile of CRP to the bottom quintile, the OR for CHD was 1.23 (0.66-2.32). The p-value for trend was 0.08. In the analyses on men, the OR for CHD was 2.55 (1.40-4.65) comparing the top quintile of CRP to the bottom quintile, and the OR for CHD was 2.14 (1.18-3.88) comparing the fourth quintile of CRP to the bottom quintile. Comparing the third quintile of CRP to the bottom quintile gave an OR for CHD of 1.74 (0.96-3.15), and comparing the second quintile of CRP to the bottom quintile gave an OR for CHD of 1.75 (0.97-3.16). The p-value for trend was 0.02. Analyses were also performed by categorising levels of CRP as low risk (< 1.0 mg/L), medium risk (1.0-2.9 mg/L), and high risk (≥ 3.0 mg/L). These analyses were adjusted for the same factors as the analysis for quintiles of CRP described above. For women, the OR for CHD was 1.53 (0.89-2.62) comparing high risk levels of CRP to low risk, and the OR for CHD was 1.17 (0.69-2.00) comparing medium risk to low risk. The p-value for trend was 0.09. For men, the OR for CHD was 1.79 (1.14-2.83) comparing high risk levels of CRP to low risk, and the OR for CHD was 1.60 (1.09-2.34) comparing medium risk to low risk. The p-value for trend was 0.03. In the analysis combining men and women, the OR for CHD was 1.68 (1.18-2.38) comparing high risk to low risk, and the OR for CHD was 1.44 (1.05-1.96) comparing medium risk to low risk.

The estimated associations described in this section are summarised in Table 6.2.

6.2.3. Case-cohort studies

Case-cohort studies are studies in which disease cases from an existing cohort study or randomised trial are compared with a random sample of the population at baseline. An advantage of this study design is that the same random sample of the study population can be used for analyses of several different outcomes.

Folsom, Aleksic, Catellier *et al.*[167] used data from the Atherosclerosis Risk In Communities (ARIC) study to investigate the association between CRP and incident CHD. A total of 615 cases occurring between 1987 and 1995 were used. In analyses adjusted for age, sex, ethnicity, centre, smoking, total cholesterol, HDL cholesterol, diabetes, systolic blood pressure, use of antihypertensives, fibrinogen, white blood cell count, and soluble intracellular adhesion molecule-1, the adjusted hazard ratio for CHD comparing the top quintile of CRP with the bottom quintile was 1.4 (0.7-2.8). Comparing the fourth quintile of CRP with the bottom quintile gave a hazard ratio for CHD of 2.1 (1.1-3.9), and comparing the third quintile of CRP to the bottom quintile gave a hazard ratio for CHD of 1.6 (0.9-3.0). The hazard ratio for CHD was 0.8 (0.5-1.6) when comparing the second quintile of CRP with the bottom quintile. The p-value for trend was 0.06.

ARIC study data was also used by Ballantyne, Hoogeveen, Bang *et al.*[168] in a case-cohort design

Table 6.3: Summary of estimated associations between C-reactive protein and coronary heart disease or cardiovascular disease from case-cohort studies.

Reference	Number	Outcome	Exposure	Control for confounding	Hazard ratio	95% CI	P value for trend
Folsom *et al.*[167]	615 cases	Incident CHD	0.83-2.07 mg/L vs. < 0.83 mg/L	ABCFG	0.8	(0.5-1.6)	0.06
			2.08-3.85 mg/L vs. < 0.83 mg/L	ABCFG	1.6	(0.9-3.0)	
			3.86-6.09 mg/L vs. < 0.83 mg/L	ABCFG	2.1	(1.1-3.9)	
			> 6.09 mg/L vs. < 0.83 mg/L	ABCFG	1.4	(0.7-2.8)	
Ballantyne *et al.*[168]	608 cases	Incident CHD	1.0-3.0 mg/L vs. < 1.0 mg/L	ABC	1.31	(0.96-1.80)	
	740 non-cases		> 3.0 mg/L vs. < 1.0 mg/L	ABC	1.72	(1.24-2.39)	

Abbreviations: CI, confidence interval; CHD, coronary heart disease; FEV_1, forced expiratory volume in one second.
A: Adjusted for general factors, such as age and sex.
B: Adjusted for behavioural risk factors, such as smoking and alcohol intake.
C: Adjusted for physiological risk factors, such as blood pressure and cholesterol.
D: Adjusted for anthropometric and/or lung function variables, such as height and FEV_1.
E: Adjusted for socio-economic status.
F: Adjusted for medications, such as statins or antihypertensive drugs.
G: Adjusted for other inflammatory markers, such as fibrinogen or interleukin-6.

to investigate the association between CRP and incident CHD. The analysis included 608 cases and 740 non-cases. Levels of CRP were defined as high risk (>3.0 mg/L), medium risk (1.0-3.0 mg/L) and low risk (<1 mg/L). Controlling for age, sex, ethnicity, smoking, systolic blood pressure, LDL cholesterol, HDL cholesterol and diabetes, the hazard ratio for incident CHD was 1.72 (1.24-2.39) comparing high risk CRP levels with low risk. Comparing medium risk CRP levels to low risk, the hazard ratio for CHD was 1.31 (0.96-1.80).

The estimated associations between CRP and CHD described in this section are summarised in Table 6.3.

6.2.4. Prospective cohort studies

Data from the MONICA project in Augsburg, Germany, were used by Koenig, Sund, Fröhlich *et al.*[169] to perform a prospective analysis of the effect of CRP on fatal and non-fatal MI. The analysis included 936 healthy men. A change in estimate criterion, in which a variable was included in the analysis if it changed the hazard ratio by 5% or more, was used to identify potential confounders. The hazard ratio for MI for a standard deviation increase in log CRP, and controlled for age and smoking, was 1.50 (1.14-1.97). This estimate is likely to be biased by unmeasured confounding. As discussed in Chapter 3, the change in estimate criterion is not a reliable way to eliminate unmeasured confounding.

Strandberg and Tilvis[170] used data on 455 people aged either 75, 80 or 85 from the Helsinki Aging Study to investigate the association between CRP and ten year cardiovascular mortality. Adjusting for age and sex, the hazard ratio for ten year cardiovascular mortality was 1.22 (1.10-1.35) per 10 mg/L increase of CRP.

Data from the Caerphilly Prospective Heart Disease Study was used by Mendall, Strachan, Butland *et al.*[171] to estimate the association between CRP and ischaemic heart disease (IHD). Analyses included 1,395 men, and adjusted for age, BMI, height, FEV_1, alcohol, smoking, current social class, father's social class, systolic blood pressure, total cholesterol and fibrinogen. Comparing the top quintile of CRP to the bottom quintile, the OR for IHD was 0.96 (0.50-1.86). The OR for IHD was 1.14 (0.60-2.15) comparing the fourth quintile of CRP to the bottom quintile, and the OR for IHD was 0.92 (0.48-1.75) comparing the third quintile of CRP to the bottom quintile. Comparing the second quintile of CRP to the bottom quintile gave an OR for IHD of 1.11 (0.58-2.10). The p-value for trend was 0.8308. The association between CRP and fatal IHD was also investigated. Comparing the top quintile of CRP to the bottom quintile gave an OR for fatal IHD of 0.79 (0.34-1.84). The OR for fatal IHD was 1.20 (0.53-2.75) comparing the fourth quintile of CRP to the bottom quintile, and the OR for fatal IHD was 0.93 (0.39-2.19) comparing the third quintile of CRP to the bottom quintile. Comparing the second quintile of CRP to the bottom quintile gave an OR for fatal IHD of 0.72 (0.29-1.76). The p-value for trend

was 0.8430.

Lowe, Yarnell, Rumley *et al.*[172] used data from the Speedwell study to investigate the association between CRP and IHD. Analyses were adjusted for age, whether the blood sample thawed during a freezer failure or not, smoking, BMI, diastolic blood pressure, total cholesterol, evidence of ischaemia at baseline, fibrinogen and fibrin D-dimer. In an analysis of 1,595 men with complete data on all variables, the OR for IHD was 1.45 (0.79-2.66) comparing the top quintile of CRP with the bottom quintile. The p-value for trend was 0.16. The ORs for IHD comparing other quintiles of CRP with the bottom quintile were not presented.

Data from the WHS were used by Ridker, Rifai, Rose *et al.*[173] to estimate the association between CRP and CVD. Analyses included 27,937 women, and were adjusted for age, treatment assignment, smoking, diabetes, blood pressure and use of HRT. The hazard ratio for CVD was 2.3 (1.6-3.4) comparing the top quintile of CRP to the bottom quintile. Comparing the fourth quintile of CRP to the bottom quintile gave a hazard ratio for CVD of 2.0 (1.3-3.0), and comparing the third quintile of CRP to the bottom quintile gave a hazard ratio for CVD of 1.6 (1.1-2.4). The hazard ratio for CVD was 1.4 (0.9-2.2) when comparing the second quintile of CRP to the bottom quintile. The p-value for trend was less than 0.001.

Blake, Rifai, Buring *et al.*[174] also estimated an association between CRP and CVD using data from the WHS. Their analysis included 15,215 women who were not using HRT at baseline, and adjusted for age, random allocation to aspirin and vitamin E, BMI, smoking, LDL cholesterol, HDL cholesterol, diabetes and blood pressure. The hazard ratio for CVD was 1.44 (p-value=0.005) when comparing CRP levels $\geq$ 3 mg/L with CRP levels < 3 mg/L.

Koenig, Sund, Fröhlich *et al.*[175] considered the impact of within-subject variation of CRP on the estimated association between CRP and CHD using data from the MONICA Augsburg Studies. Analyses included 936 subjects. The ICC of CRP was estimated as 0.54, and a Bayesian method was used to obtain the estimates corrected for within-subject variation. Adjusting for age and BMI, the uncorrected hazard ratio for CHD was 1.55 (1.23-1.95) for a unit increase in log CRP. Allowing for within-subject variation in CRP, the corrected hazard ratio for CHD was 2.59 (1.61-4.16) for a unit increase in log CRP.

Data from the Health, Aging and Body Composition (Health ABC) Study was used by Cesari, Penninx, Newman *et al.*[176] to investigate the association between CRP and CVD. The study population consisted of 3,045 participants, aged between 70 and 79 years at baseline. Analyses were adjusted for age, sex, ethnicity, education, smoking, diabetes, hypertension, cancer, BMI, total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides, creatinine, non-steroidal anti-inflammatory drugs, angiotensin-converting enzyme inhibitors and statin use. Participants

with subclinical CVD were excluded from analyses with clinical CVD as the outcome. The OR for clinical CVD was 1.08 (0.96-1.21) per log μg/ml increase in CRP. Comparing the top tertile of CRP with the bottom tertile gave an OR for clinical CVD of 1.34 (1.03-1.65), and comparing the middle tertile of CRP with the bottom tertile gave an OR for clinical CVD of 1.28 (0.99-1.64).

In a separate publication, Cesari, Penninx, Newman *et al.*[177] again used data from the Health ABC Study to investigate the association between CRP and CHD. Their analysis included 2,225 subjects aged between 70 and 79 at baseline. Analyses adjusted for age, sex, ethnicity, smoking, diabetes, hypertension, BMI, HDL cholesterol, triglycerides and albumin. The hazard ratio for CHD was 1.13 (0.95-1.35) for a unit increase in log CRP. For a standard deviation increase in log CRP, the hazard ratio for CHD was 1.11 (0.96-1.29). Analyses were also conducted in which CRP was divided into tertiles. The hazard ratio for CHD was 1.09 (0.76-1.57) comparing the middle tertile of CRP to the bottom tertile, and the hazard ratio for CHD was 1.20 (0.83-1.75) comparing the top tertile of CRP to the bottom tertile.

Lowe, Sweetnam, Yarnell *et al.*[178] investigated the association between CRP and IHD using data from the Caerphilly and Speedwell prospective cohort studies. The multivariable analysis included 3,065 men, and adjusted for age, area, smoking, BMI, diastolic blood pressure, total cholesterol, evidence of ischaemia at baseline, fibrinogen and fibrin D-dimer. The OR for IHD was 1.51 (0.98-2.33) comparing the top quintile of CRP to the bottom quintile. The p-value for trend was 0.093. The ORs for IHD comparing lower quintiles of CRP with the bottom quintile were not presented.

Data from the Strong Heart Study was used by Best, Zhang, Lee *et al.*[179]. The study cohort consisted of 3,277 subjects without CVD at baseline. A stepwise selection method was used to identify the confounders of the association between CRP and CVD. This resulted in adjustment for sex, age, smoking, LDL cholesterol, HDL cholesterol, hypertension and diabetes. The hazard ratio for CVD was 1.25 (0.90-1.75) when comparing the top quintile of CRP to the bottom quintile. Comparing the fourth quintile of CRP to the bottom quintile gave a hazard ratio for CVD of 1.31 (0.95-1.81), and comparing the third quintile of CRP to the bottom quintile gave a hazard ratio for CVD of 1.17 (0.84-1.63). The hazard ratio for CVD was 1.0 (0.72-1.40) when comparing the second quintile of CRP to the bottom quintile. Analyses were also conducted in which subjects with baseline levels of CRP greater than 10 mg/L were excluded, leaving 2,735 subjects for the analysis. The hazard ratio for CVD was 1.63 (1.00-2.67) comparing high risk levels of CRP (> 3 mg/L) to low risk (< 1 mg/L). Comparing medium risk levels of CRP (1-3 mg/L) to low risk, the hazard ratio for CVD was 1.35 (0.82-2.22).

St-Pierre, Cantin, Bergeron *et al.*[180] used data from the Québec Cardiovascular Study to investigate the association between CRP and IHD. Analyses included 1,982 men who did not

have IHD at baseline. In an analysis adjusted for age, BMI, systolic blood pressure, diabetes, medication use, LDL cholesterol, HDL cholesterol, triglycerides, fibrinogen and interleukin-6, the hazard ratio for incident IHD was 0.70 (0.43-1.13) when comparing the top quartile of CRP with the bottom quartile. Hazard ratios for IHD when comparing the second or third quartile of CRP with the bottom quartile were not given.

Data from the WHS was used by Ridker, Rifai, Cook *et al.*[181] to investigate the association between CRP and ten year incidence of CVD among women not using HRT at baseline. The analyses included 15,532 women, and were adjusted for age, blood pressure, BMI, diabetes and smoking. The hazard ratio for CVD was 2.98 (1.90-4.67) comparing the top quintile of CRP to the bottom quintile. Comparing the fourth quintile of CRP to the bottom quintile gave a hazard ratio for CVD of 2.38 (1.52-3.72), and comparing the third quintile of CRP to the bottom quintile gave a hazard ratio for CVD of 1.91 (1.21-3.03). The hazard ratio for CVD was 1.85 (1.16-2.96) comparing the second quintile of CRP to the bottom quintile. The p-value for trend was less than 0.001.

Laaksonen, Niskanen, Nyyssönen *et al.*[182] used data from the Kuopio Ischaemic Heart Disease Risk Factor Study to investigate the association between CRP and CVD death. The analyses included 2,321 men without diabetes or cancer at baseline. Analyses were adjusted for age, year of examination, LDL cholesterol, systolic blood pressure, use of blood pressure medication, cigarette smoking, dietary variables, insulin resistance, exercise, alcohol intake and socioeconomic status. Classifying people's CRP levels into low risk (< 1 mg/L), medium risk (1-<3 mg/L) and high risk (≥ 3 mg/L), the hazard ratio for CVD death was 2.90 (1.36-6.19) comparing high risk to low risk among men without CVD at baseline. The hazard ratio for CVD death was 2.39 (1.29-4.44) comparing medium risk to low risk among men without CVD at baseline. The p-value for trend was 0.004. Considering people with CVD at baseline, the hazard ratio for CVD death was 1.91 (1.12-3.27) comparing high risk levels of CRP to low risk, and the hazard ratio for CVD death was 1.15 (0.69-1.93) when comparing medium risk to low risk. The p-value for trend for this analysis was 0.010. Hazard ratios were also estimated for comparisons between tertiles of CRP. Including only subjects without CVD at baseline in the analysis, the hazard ratio for CVD death was 3.88 (1.72-8.73) comparing the top tertile of CRP to the bottom tertile, and the hazard ratio for CVD death was 3.65 (1.65-8.05) comparing the middle tertile of CRP with the bottom tertile. The p-value for trend was 0.002. Considering men with CVD at baseline, the hazard ratio for CVD death was 1.73 (0.99-3.04) comparing the top tertile of CRP to the bottom tertile. Comparing the middle tertile of CRP to the bottom tertile gave a hazard ratio for CVD death of 1.23 (0.69-2.20). The p-value for trend was 0.032.

Data from the CHS was used by Cushman, Arnold, Psaty *et al.*[183] to investigate the association between CRP and ten year incidence of CHD. Their analysis was based on 3,971 participants

Table 6.4: Summary of estimated associations between C-reactive protein and coronary heart disease or cardiovascular disease from prospective cohort studies.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Koenig *et al.*[169]	936 men	Incident CHD	Standard deviation increase in log CRP	AB	1.50	(1.14-1.97)	
Strandberg *et al.*[170]	455 elderly people	Fatal CVD	10 mg/L increase	A	1.22[a]	(1.10-1.35)	
Lowe *et al.*[172]	1,595 men	Incident IHD	Top quintile vs. bottom quintile	ABCG	1.45	(0.79-2.66)	0.16
Mendall *et al.*[171]	1,395 men	Incident IHD	Second quintile vs. bottom quintile	ABCDEG	1.11	(0.58-2.10)	0.8308
			Third quintile vs. bottom quintile	ABCDEG	0.92	(0.48-1.75)	
			Fourth quintile vs. bottom quintile	ABCDEG	1.14	(0.60-2.15)	
			Top quintile vs. bottom quintile	ABCDEG	0.96	(0.50-1.86)	
		Fatal IHD	Second quintile vs. bottom quintile	ABCDEG	0.72	(0.29-1.76)	0.8430
			Third quintile vs. bottom quintile	ABCDEG	0.93	(0.39-2.19)	
			Fourth quintile vs. bottom quintile	ABCDEG	1.20	(0.53-2.75)	
			Top quintile vs. bottom quintile	ABCDEG	0.79	(0.34-1.84)	
Ridker *et al.*[173]	27,939 women	Incident CVD	> 0.49-1.08 mg/L vs. ≤ 0.49 mg/L	ABCF	1.4[a]	(0.9-2.2)	<0.001
			> 1.08-2.09 mg/L vs. ≤ 0.49 mg/L	ABCF	1.6[a]	(1.1-2.4)	
			> 2.09 mg/L vs. ≤ 0.49 mg/L	ABCF	2.0[a]	(1.3-3.0)	
			> 4.19 mg/L vs. ≤ 0.49 mg/L	ABCF	2.3[a]	(1.6-3.4)	
Blake *et al.*[174]	15,215 women (not using HRT)	Incident CVD	≥ 3 mg/L vs. < 3 mg/L	ABC	1.44[a]	p=0.005	
Koenig *et al.*[175]	936 subjects	Incident CHD	Unit increase in log CRP (adjusted for within-subject variation in CRP)	AB	2.59[a]	(1.61-4.16)	

Table 6.4 continued.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Cesari *et al.*[176]	3,045 elderly people	Incident CVD	Unit increase in log CRP	ABCF	1.08	(0.96-1.21)	
		Incident CVD	Middle tertile vs. bottom tertile	ABCF	1.28	(0.99-1.64)	
			Top vs. bottom	ABCF	1.34	(1.03-1.65)	
Cesari *et al.*[177]	2,225 subjects	Incident CHD	Unit increase in log CRP	ABC	1.13[a]	(0.95-1.35)	
		Incident CHD	Standard deviation increase in log CRP	ABC	1.11[a]	(0.96-1.29)	
		Incident CHD	1.16-2.50 mg/L vs. 0.15-1.15 mg/L	ABC	1.09[a]	(0.76-1.57)	
			2.51-85.18 mg/L vs. 0.15-1.15 mg/L	ABC	1.20[a]	(0.83-1.75)	
Lowe *et al.*[178]	3,065 men	Incident IHD	Top quintile vs. bottom quintile	ABCG	1.51	(0.98-2.33)	0.093
Best *et al.*[179]	3,277 subjects	Incident CVD	1.7-3.1 mg/L vs. < 1.7 mg/L	ABC	1.0[a]	(0.72-1.40)	
			3.1-4.8 mg/L vs. < 1.7 mg/L	ABC	1.17[a]	(0.84-1.63)	
			4.8-8.7 mg/L vs. < 1.7 mg/L	ABC	1.31[a]	(0.95-1.81)	
			8.7-123.4 mg/L vs. < 1.7 mg/L	ABC	1.25[a]	(0.90-1.75)	
	2,735 subjects (with baseline CRP <= 10 mg/L)	Incident CVD	1-3 mg/L vs. < 1 mg/L	ABC	1.35[a]	(0.82-2.22)	
			> 3 mg/L vs. < 1 mg/L	ABC	1.63[a]	(1.00-2.67)	
St-Pierre *et al.*[180]	1,982 men (without IHD at baseline)	Incident IHD	≥ 3.80 mg/L vs. < 0.84 mg/L	ABCFG	0.70	(0.43-1.13)	
Ridker *et al.*[181]	15,632 women (not using HRT)	Incident CVD	Second quintile vs. bottom quintile	ABC	1.85[a]	(1.16-2.96)	<0.001
			Third quintile vs. bottom quintile	ABC	1.91[a]	(1.21-3.03)	
			Fourth quintile vs. bottom quintile	ABC	2.38[a]	(1.52-3.72)	
			Top quintile vs. bottom quintile	ABC	2.98[a]	(1.90-4.67)	

Table 6.4 continued.

Reference	Number	Outcome	Exposure	Control for confounding	Odds ratio	95% CI	P-value for trend
Laaksonen *et al.*[182]	1,476 men (without CVD at baseline)	Fatal CVD	1.00-2.99 mg/L vs. 0.10-0.99 mg/L	ABCEF	2.39[a]	(1.29-4.44)	0.004
			3.00-9.99 mg/L vs. 0.10-0.99 mg/L	ABCEF	2.90[a]	(1.36-6.19)	
	845 men (with CVD at baseline)	Fatal CVD	1.00-2.99 mg/L vs. 0.10-0.99 mg/L	ABCEF	1.15[a]	(0.69-1.93)	0.010
			3.00-9.99 mg/L vs. 0.10-0.99 mg/L	ABCEF	1.91[a]	(1.12-3.27)	
	1,476 men (without CVD at baseline)	Fatal CVD	0.84-1.82 mg/L vs. 0.10-0.83 mg/L	ABCEF	3.65[a]	(1.65-8.05)	0.002
			1.83-9.99 mg/L vs. 0.10-0.83 mg/L	ABCEF	3.88[a]	(1.72-8.73)	
	845 men (with CVD at baseline)	Fatal CVD	0.84-1.82 mg/L vs. 0.10-0.83 mg/L	ABCEF	1.23[a]	(0.69-2.20)	0.032
			1.83-9.99 mg/L vs. 0.10-0.83 mg/L	ABCEF	1.73[a]	(0.99-3.04)	
Cushman *et al.*[183]	3,971 subjects (age ≥ 65)	Incident CHD	1.0-3.0 mg/L vs. < 1.0 mg/L	ABCDF	1.04[a]	(0.82-1.31)	
			> 3.0 mg/L vs. < 1.0 mg/L	ABCDF	1.37[a]	(1.06-1.78)	
Lawlor *et al.*[153]	2,723 women (without CHD at baseline)	Incident CHD	Doubling of CRP	ABCDE	1.03[a]	(0.94-1.13)	

Abbreviations: CI, confidence interval; CHD, coronary heart disease; CVD, cardiovascular disease; IHD, ischaemic heart disease; CRP, C-reactive protein; HRT, hormone replacement therapy; FEV_1, forced expiratory volume in one second.

[a] *Hazard ratio*

A: Adjusted for general factors, such as age and sex.

B: Adjusted for behavioural risk factors, such as smoking and alcohol intake.

C: Adjusted for physiological risk factors, such as blood pressure and cholesterol.

D: Adjusted for anthropometric and/or lung function variables, such as height and FEV_1.

E: Adjusted for socio-economic status.

F: Adjusted for medications, such as statins or antihypertensive drugs.

G: Adjusted for other inflammatory markers, such as fibrinogen or interleukin-6.

without vascular diseases at baseline and aged 65 or over at baseline. Levels of CRP were defined as high risk (> 3.0 mg/L), medium risk (1.0-3.0 mg/L) or low risk (< 1.0 mg/L). Analyses were adjusted for age, sex, race, field centre, hypertension, diabetes, smoking, log pack-years, BMI, waist circumference, total cholesterol, HDL cholesterol, regular aspirin use, and measures of subclinical vascular disease. The hazard ratio for CHD was 1.37 (1.06-1.78) comparing high risk CRP levels with low risk, and the hazard ratio for CHD was 1.04 (0.82-1.31) comparing medium risk CRP levels with low risk.

As well as the cross-sectional analysis described above in Section 6.2.1, Lawlor *et al.*[153] investigated the effect of doubling CRP on time to incident CHD among women without prevalent CHD at baseline. Data for 2,723 women were used in the analysis. Upon adjustment for age, life course socio-economic position score, behavioural and physiological risk factors, adult anthropometry and FEV_1, the hazard ratio for incident CHD for a doubling of CRP was 1.03 (0.94-1.13).

The estimated associations between CRP and CHD or CVD described in this section are summarised in Table 6.4.

6.2.5. Meta-analyses

A meta-analysis of prospective studies to estimate the effect of blood levels of CRP on risk of CHD was performed by Danesh, Collins, Appleby *et al.*[184] Seven prospective studies were identified and included in the meta-analysis. The risk ratio for CHD was found to be 1.7 (1.4-2.1), and corresponded to a difference in CRP of 1.4 mg/L. This meta-analysis was updated in 2000 by Danesh, Whincup, Walker *et al.*[160] to include an additional seven prospective studies. The overall risk ratio for CHD was found to be 1.9 (1.5-2.3) when comparing the top third with the bottom third of CRP, and corresponded to a difference of CRP of 1.4 mg/L. A further update was carried out in 2004 and included 22 prospective studies.[165] The OR for CHD, comparing the top with the bottom tertile of the CRP distribution, was 1.58 (1.48-1.68).

6.3. Summary

Many studies have shown a positive association between CRP and CHD. There was, however, very little consideration of possible bias due to measurement error in the exposure, residual confounding or unmeasured confounding in these studies. Only one observational study was identified which investigated the effects of measurement error, and this study only considered measurement error in CRP.[175] Their conclusion was that the estimated association between CRP and CHD, based on a single CRP measurement, would be an underestimate of the true association. Their analysis only adjusted for confounding by age and BMI, both of which were considered to be perfectly measured. It seems likely that their estimated association is biased by unmeasured confounding. Naïve estimates of the association between CRP and CHD, in

which there is no correction for the effects of measurement error, may not underestimate the association if the effects of any residual and unmeasured confounding act to bias the estimated association away from the null value. It is also possible that the observed associations are a result of reverse causation, where elevated CRP levels are a result of pre-clinical disease rather than a cause. It therefore remains open to discussion whether the estimated associations between CRP and CHD from observational epidemiological studies are causal.

Davey Smith, Harbord and Ebrahim[185] suggested that Mendelian randomisation could address these questions. In parent-offspring designs, the genetic variant associated with raised CRP levels does not suffer from confounding, as a person's genetic status is determined by segregation of genes, which is a random process. Mendelian randomisation in other designs for genetic association studies will be approximate.[186] Additionally, the genes are determined before birth, which eliminates the possibility of reverse causation. One drawback of this method is that large sample sizes are required to give a sufficiently precise estimate. In the study by Zee and Ridker,[187] no association was found between the genetic variant associated with elevated CRP levels and CHD. A similar lack of association was found in a genetic study by Casas, Shah, Cooper *et al.*,[188] which suggests that the estimated associations from observational epidemiological studies may not be causal.

A further hypothesis is that confounding by factors acting across the lifecourse could explain the observed CRP/CHD association.[153] In a study using data from the BWHHS, Lawlor, Davey Smith, Rumley *et al.*[153] showed that the inclusion of variables describing socio-economic status across the lifecourse, behavioural and physiological risk factors, adult anthropometry and lung function attenuated the observed association between CRP and both prevalent CHD at baseline and incident CHD.

The effects of exposure measurement error and residual confounding on the estimated association between CRP and CHD are investigated in Chapter 7.

BLANK IN ORIGINAL

Chapter 7.

Case study: Allowing for measurement error in confounders in estimating the association between C-reactive protein and coronary heart disease

7.1. Introduction

In this chapter the impact of residual confounding and exposure measurement error on the estimated association between C-reactive protein (CRP) and coronary heart disease (CHD) is investigated, using data from the British Women's Heart and Health Study (BWHHS). The effects of residual confounding by forced expiratory volume in one second (FEV_1) and triglycerides are investigated. These were chosen based on their crude and multivariable estimated associations with CHD. A Monte Carlo sensitivity analysis, described in Section 4.4.3, is used as the BWHHS data do not contain validation or replication data on CRP, FEV_1 or triglycerides. This Monte Carlo sensitivity analysis involves repeatedly choosing measurement error variances for the two confounders and the exposure from three probability distributions. Repeatability measures obtained from the literature, such as intra-class correlation coefficients (ICCs) and coefficients of variation (CVs), are used to define these probability distributions. Regression calibration is then used to correct estimated associations for measurement error in the confounders and exposure. The analysis is carried out for three CHD outcomes; prevalent CHD, incident CHD, and time to CHD.

7.2. Methods

7.2.1. Notation and terminology

The notation introduced in Chapter 5 (Section 5.2.1) will also be used in this chapter. Throughout this chapter the term *naïve analysis* will refer to the analysis in which no adjustment is made for measurement error, and the term *crude analysis* will refer to any analysis in which no control for confounders is made.

7.2.2. The British Women's Heart and Health Study

The BWHHS is a prospective cohort study. The selection of participants and measurements used have been described previously by Lawlor, Ebrahim and Davey Smith[189] and Lawlor, Bedford, Taylor, *et al.*,[190] so will be only briefly reviewed here. Between 1999 and 2001, 4,286 women aged 60 to 79 years, who were randomly selected from 23 British towns, were interviewed, examined, completed medical questionnaires and had detailed reviews of their medical records. These women have been followed up by flagging with the NHS central register for mortality data, two yearly review of their medical records and a 3-year follow-up questionnaire sent to all surviving participants between March and September 2003. Local ethics committees' approvals were obtained for the study.

The BWHHS data contain information on prevalent CHD at baseline, incidence of CHD, and time to incident CHD. Prevalent CHD at baseline was defined as a woman with either a

medical record of myocardial infarction, angina, coronary artery bypass or angioplasty, or self-report that a doctor had ever diagnosed a heart attack or angina. Incident cases of CHD in women who were free of prevalent CHD at baseline were defined as death with an underlying cause of CHD, a new myocardial infarction, a new diagnosis of angina or a new coronary artery bypass or angioplasty identified by medical record review, or self-report of a new heart attack or new diagnosis of angina. The analyses in this chapter will investigate each outcome (prevalent CHD, incident CHD and time to incident CHD), using a logistic regression model for presence of prevalent and incident CHD, and a Cox proportional hazards model for time to incident CHD. The confounders that will be adjusted for in all analyses are age, smoking status (never smoked, ex-smoker, or current smoker), systolic blood pressure, total cholesterol, triglycerides (logged), body mass index, FEV_1, presence of diabetes, and socio-economic status in six categories (class I, class II, class III non-manual, class III manual, class IV, and class V). These confounders are commonly adjusted for in other published analyses.

Measurement error in two confounders, log triglycerides and FEV_1, and in CRP will be considered. The reasons for choosing these confounders will be presented later (see Section 7.3.1). In order to correct for measurement error in continuous explanatory variables, the relevant measurement error variance should be estimated from repeated measures or an internal or external validation study. This information is not available in the BWHHS data, and therefore repeatability measures for CRP, FEV_1 and triglycerides, such as ICCs or CVs, were extracted from the literature. To find articles on FEV_1, Medline was searched for articles including the terms FEV_1, spirometry, spirometric, or forced expiratory in the title. In the case of triglycerides, articles were searched for the terms triglyceride or lipid in the title. For CRP, titles were searched for the terms C-reactive protein, CRP, inflammation or inflammatory. These were combined with searches on the terms measurement error, reliability, reproducibility, repeatability, variability or variation. Only articles in English were considered. After the searches, titles and abstracts were reviewed and any relevant articles obtained. Only papers which measured repeatability on a healthy population, or on a randomly selected population were included. Additionally, papers in which repeated measures were obtained on the same day were excluded. The reference lists of the extracted articles were also reviewed, and any additional relevant papers obtained.

7.2.3. Sensitivity analysis method

To investigate the impact of residual confounding by FEV_1 and triglycerides and measurement error in CRP on the estimated association between CRP and CHD, a Monte Carlo risk analysis method will be used (see Section 4.4.3). Most of the applications of Monte Carlo risk analysis to measurement error in the literature consider only misclassification of categorical variables (e.g. Fox *et al.*,[134] Phillips,[136] Lash *et al.*,[61] Lash *et al.*[135]). As the analysis of interest in this chapter investigates the effect of measurement errors in continuous variables, the exact methods used in

these papers cannot be used for this analysis. A Monte Carlo risk analysis was therefore developed to investigate the effect of measurement error in continuous variables.

The first step in the Monte Carlo method is to independently choose values of the error variances for each of the variables (FEV_1, log triglycerides and CRP) from probability distributions. The method for choosing the probability distributions for each of the variables is described in detail below.

7.2.3.1 Choosing a probability distribution for the error variance of FEV_1

First, the method of choosing a probability distribution for the error variance of FEV_1 is described. This probability distribution will be informed by repeatability measures found in the literature, such as ICCs or CVs. As FEV_1 is approximately normally distributed in the BWHHS data, a classical measurement error structure in the confounders is assumed, where the measurement error is additive and independent of the true values. Values of $(\sigma_u^2 + \sigma_e^2)$ and μ can therefore be observed from the data. The formulae for deriving the measurement error variances for the BWHHS from the ICC or CV values in the literature are therefore

Equation 7.1: $$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \quad \Rightarrow \quad \sigma_e^2 = (\sigma_u^2 + \sigma_e^2) - ICC(\sigma_u^2 + \sigma_e^2),$$

Equation 7.2: $$CV = \frac{\sigma_e}{\mu} \quad \Rightarrow \quad \sigma_e^2 = (CV\mu)^2.$$

A normal probability distribution is then chosen so that 95% of error variances chosen will be in the range found in the literature. The mean of the normal distribution is defined to be the mid-point between the largest and smallest derived error variance. If $\sigma^2_{e,max}$ and $\sigma^2_{e,min}$ are the largest and smallest derived error variances respectively, the mean of the normal distribution is therefore equal to

$$\sigma^2_{e,min} + \frac{(\sigma^2_{e,max} - \sigma^2_{e,min})}{2}.$$

The variance for the normal distribution is also defined using the largest and smallest derived error variances. From the properties of a normal distribution, values 1.96 standard deviations either side of the mean occur with probability 0.95. The variance of the probability distribution is therefore defined so that the values between the smallest and largest derived error variances are chosen with probability 0.95. This is equivalent to setting the range between the smallest and largest derived error variances equal to 3.92 standard deviations of the required normal distribution. The variance is therefore equal to

$$\left(\frac{\sigma^2_{e,max} - \sigma^2_{e,min}}{3.92}\right)^2.$$

As error variances cannot be negative, this normal distribution is truncated at zero. In practice, truncating the chosen distribution at zero makes very little difference to the coverage probabilities. The probability of choosing negative error variances, and therefore the impact of

truncating the distributions at zero, for the normal distributions used in this analysis are provided in Section 7.3.2.

7.2.3.2 Choosing a probability distribution for the error variance of log triglycerides

The method used to choose the probability distribution for the error variance of logged triglycerides is slightly different to that described for FEV_1 in Section 7.2.3.1. This difference arises because the triglyceride variable is skewed in the BWHHS data, and the logged value will be used in the regression analyses presented later (Section 7.3). The repeatability measures found in the literature in general are for the skewed triglyceride variable, and therefore the following transformation was developed to derive error variances for logged triglycerides. For these transformations, it is assumed that triglycerides have a log-normal distribution, and hence that log triglycerides has a normal distribution.

To calculate an error variance for log triglycerides from an ICC of triglycerides, it is first assumed that the ICC is the ratio of the variance of the true variable to the observed variance. The variance of a log-normal variable is given by $\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$, where σ^2 and μ are the variance and mean of the normally distributed variable respectively. The ICC of a log-normal variable is therefore given by

$$ICC=\frac{\left(e^{\sigma_x^2}-1\right)e^{2\mu_x+\sigma_x^2}}{\left(e^{\sigma_x^2+\sigma_e^2}-1\right)e^{2\mu_x+\sigma_x^2+\sigma_e^2}}.$$

As μ_X and $\left(\sigma_X^2+\sigma_e^2\right)$, the mean and variance of the observed log triglycerides variable, are observable from the data, $e^{\sigma_x^2}$ can be calculated from the above formula.

$$0=e^{2\mu_x+2\sigma_x^2}-e^{2\mu_x+\sigma_x^2}-ICC\left(e^{\sigma_x^2+\sigma_e^2}-1\right)e^{2\mu_x+\sigma_x^2+\sigma_e^2}$$

$$e^{\sigma_x^2}=\frac{e^{2\mu_x}\pm\sqrt{e^{4\mu_x}+4ICCe^{4\mu_x}\left(e^{\sigma_x^2+\sigma_e^2}-1\right)e^{\sigma_x^2+\sigma_e^2}}}{2e^{2\mu_x}}$$

$$=\frac{1}{2}\pm\frac{1}{2}\sqrt{1+4ICC\left(e^{\sigma_x^2+\sigma_e^2}-1\right)e^{\sigma_x^2+\sigma_e^2}}$$

As the above expression for $e^{\sigma_x^2}$ must be positive, the second half of the formula cannot be subtracted from the first. The error variance of log triglycerides is therefore given by

$$\sigma_e^2=\left(\sigma_X^2+\sigma_e^2\right)-\ln\left(\frac{1}{2}+\frac{1}{2}\sqrt{1+4ICC\left(e^{\sigma_x^2+\sigma_e^2}-1\right)e^{\sigma_x^2+\sigma_e^2}}\right).$$

To calculate the error variance of log triglycerides from a CV, the CV is first converted to an ICC using the following formula

$$ICC=\frac{\sigma_Z^2-(CV\mu)^2}{\sigma_Z^2},$$

where σ_Z^2 and μ are the observed variance and mean of the untransformed triglyceride variable

respectively. This expression is derived by equating Equation 7.1 and Equation 7.2.
The CV is converted to an ICC because the CV does not place any restriction on the size of the error variance compared with the observed variance. It would therefore be possible to estimate an error variance for log triglycerides that is larger than the observed variance. This problem does not occur when an ICC is used. Following conversion of the CV to an ICC, the method described previously for estimating the error variance of log triglycerides from an ICC is used.

Once error variances for log triglycerides have been derived from the repeatability measures in the literature, a normal probability distribution is defined so that 95% of the possible error variances are within the range defined by the values in the literature. The method for deriving the parameters of this distribution has been described in Section 7.2.3.1.

7.2.3.3 Choosing a probability distribution for the error variance of CRP

The CRP variable in the BWHHS data is skewed. Using a logarithmic transformation improves the normality of the variable. In the analyses, the effect of a doubling of CRP on the CHD outcome is investigated, and therefore log CRP is also divided by log 2. Dividing by a constant has no effect on ICCs or CVs. The procedure described in Section 7.2.3.2 can therefore be used to derive the error variance from ICC or CV values based on skewed CRP measurements obtained from the literature. If ICC or CV values for CRP from the literature use a logged CRP variable, the method described in Section 7.2.3.1 for deriving an error variance for FEV_1 can be used. Following derivation of the measurement error variance, a normal distribution is defined so that 95% of the error variances are in the range described by the literature. The parameters of this distribution are derived using the method described in Section 7.2.3.1.

7.2.3.4 Regression calibration

An error variance for each variable assumed to be measured with error is chosen from the relevant probability distributions. Once the error variance for each variable has been chosen, regression calibration is used to correct the naïve analyses for residual confounding in the continuous confounders and for measurement error in the exposure. The regression calibration method has been described previously in Chapter 5 and will not be repeated here. For all analyses in which errors in more than one variable are considered, the errors are assumed to be independent. This implies that the off-diagonal elements of $\hat{\Sigma}_{uu}$, the estimated covariance matrix of the errors, will be zero.

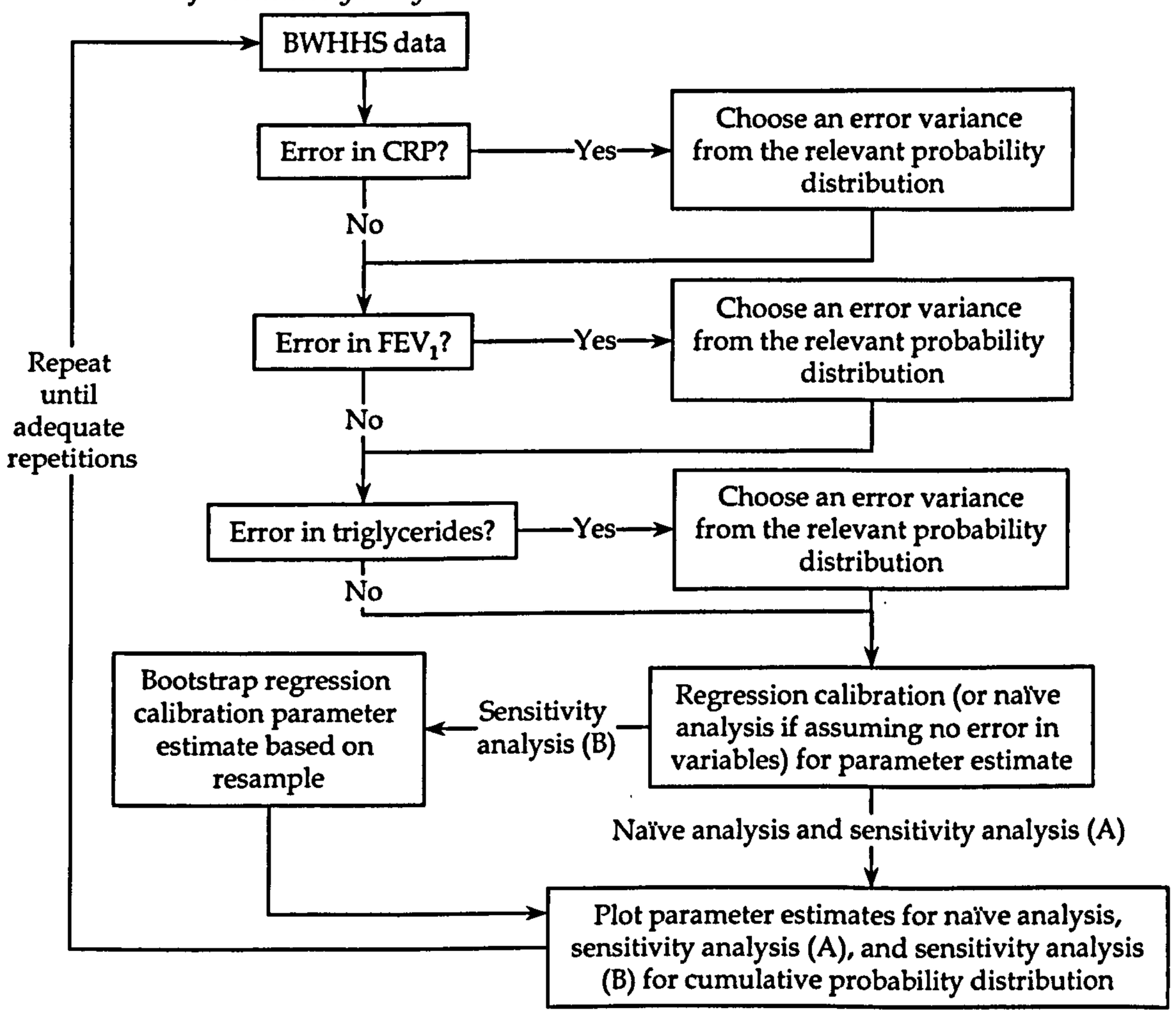
The regression calibration estimate is saved, and the process of selecting a measurement error variance and using the regression calibration method to obtain a parameter estimate is repeated until a sufficiently large number of repetitions are reached. The occurs when the widths of the 95% confidence intervals (CIs) around the 2.5th, 50th and 97.5th percentiles of the cumulative probability distribution of the parameter estimate are less than 0.01. If the stopping criteria

have not been satisfied once 100,000 repetitions have been performed, the analysis is stopped. These occurrences will be marked in the results section. The cumulative probability distribution is summarized using the 2.5th, 50th and 97.5th percentiles. The cumulative probability distribution of the estimate can also be plotted.

In the Monte Carlo sensitivity analysis described so far, the variability due to uncertainty in the true error variance is included, but sampling variation in the estimated parameter is ignored. This will be known as sensitivity analysis A. In sensitivity analysis B, sampling variation is reintroduced by taking a bootstrap sample of the full dataset. The method described for sensitivity analysis A is then followed, using the bootstrapped dataset, to obtain a cumulative probability distribution. Figure 7.1 shows a flowchart of the method.

The results from sensitivity analysis A include only variability due to differences in the repeatability measures extracted from the literature. If there is little variation in the measures found in the literature there will be very little variability in the cumulative probability distribution obtained from sensitivity analysis A. The results from sensitivity analysis B can be thought of as combining the sources of variation from the naïve analysis and sensitivity analysis A; sampling variation and uncertainty over the measurement error variance.

Figure 7.1: Method for sensitivity analysis.



Throughout this chapter, the results will be presented in terms of the median estimates, which is defined as the 50th percentile of the cumulative probability distribution from each sensitivity analysis, and a 95% simulation interval, which is defined as the interval between the 2.5th and 97.5th percentiles of the cumulative probability distribution for each sensitivity analysis. The results of a Monte Carlo sensitivity analysis do not have a frequentist interpretation, and *"the only formal statistical interpretation"* is Bayesian.[133] As such, the median estimate and 95% simulation interval provide summaries of a probability distribution for the effect estimate of interest, given the assumed distributions for the measurement error variances and the observed data.

7.3. Results

7.3.1. Naïve analysis

Table 7.1 and Table 7.2 show the crude and multivariable analyses respectively for the effects of each explanatory variable on prevalent and incident CHD and time to incident CHD. The confounders that will be investigated for the effects of measurement error are those continuous confounders for which the p-value for the estimated association between the confounder and CHD is less than or equal to 0.05 for all CHD outcomes, and in both the crude (Table 7.1) and multivariable (Table 7.2) analyses. Only one of the continuous confounders, FEV_1, meets these criteria. Triglycerides, however, are very close to meeting the criteria, with p-values in all analyses of 0.06 or less. The effect of measurement error in FEV_1 and triglycerides on the estimated association between CRP and CHD will therefore be investigated.

All analyses will be performed using data from participants with complete information on all confounders. From the original dataset with 4,286 observations, this resulted in the loss of 666 (15.5%) of the observations. The time to incident CHD analysis includes 13,089 person-years of observation, and a mean follow-up per person of 4.3 years. The shortest follow-up for this analysis was 2 days, while the longest was 5.5 years. Table 7.3: shows basic descriptions of the three variables of interest (FEV_1, triglycerides and CRP) for each of the analysis methods (prevalent CHD, incident CHD and time to incident CHD).

Table 7.1: Results of naïve crude analyses for the effect of all explanatory variables on prevalent and incident coronary heart disease, and time to incident coronary heart disease.

Variable	Prevalent CHD (Observations=3,620 CHD events=574)			Incident CHD (Observations=3,046 CHD events=159)			Time to incident CHD (Observations=3,046 CHD events=159)		
	Odds ratio	95% confidence interval	P value	Odds ratio	95% confidence interval	P value	Hazard ratio	95% confidence interval	P value
C-reactive protein	1.17	1.10-1.24	<0.001	1.16	1.05-1.28	0.004	1.14	1.03-1.26	0.008
Age	1.06	1.04-1.07	<0.001	1.04	1.01-1.08	0.003	1.18	1.04-1.33	0.008
Never smoked	1	-	-	1	-	-	1	-	-
Ex smoker	1.22	1.01-1.48	0.044	1.60	1.12-2.28	0.009	1.56	1.10-2.22	0.013
Current smoker	1.00	0.74-1.34	0.980	2.22	1.42-3.49	<0.001	2.43	1.57-3.77	<0.001
Systolic blood pressure[a]	0.91	0.83-0.99	0.032	1.24	1.06-1.46	0.007	1.19	1.01-1.40	0.035
Total cholesterol	0.77	0.71-0.83	<0.001	1.04	0.92-1.19	0.527	1.04	0.91-1.18	0.557
Triglycerides (logged)	1.55	1.28-1.87	<0.001	1.69	1.21-2.35	0.002	1.66	1.20-2.30	0.002
Body mass index	1.05	1.03-1.07	<0.001	1.01	0.98-1.05	0.399	1.02	0.98-1.05	0.337
FEV_1	0.44	0.37-0.53	<0.001	0.49	0.36-0.67	<0.001	0.52	0.37-0.73	<0.001
Diabetes	2.18	1.55-3.06	<0.001	1.22	0.59-2.55	0.593	1.21	0.60-2.47	0.594
SES Class I	1	-	-	1	-	-	1	-	-
SES Class II	1.33	0.88-2.01	0.180	1.17	0.59-2.31	0.659	1.29	0.64-2.59	0.472
SES Class III (non-manual)	1.62	1.07-2.44	0.022	1.54	0.79-3.01	0.207	1.54	0.77-3.05	0.220
SES Class III (manual)	2.27	1.49-3.48	<0.001	1.74	0.85-3.54	0.127	1.79	0.87-3.69	0.114
SES Class IV	2.34	1.55-3.55	<0.001	1.34	0.66-2.73	0.422	1.46	0.71-3.01	0.302
SES Class V	2.49	1.37-4.53	0.003	1.29	0.40-4.16	0.671	1.37	0.43-4.36	0.599

Abbreviations: CHD, coronary heart disease; FEV_1, forced expiratory volume in one second; SES, socio-economic status.

[a] Parameter estimates are per standard deviation increase (25.1 mmHg)

Table 7.2: Results of naïve multivariable analyses for the effect of all explanatory variables on prevalent and incident coronary heart disease, and time to incident coronary heart disease.

Variable	Prevalent CHD (Observations=3,620 CHD events=574)			Incident CHD (Observations=3,046 CHD events=159)			Time to incident CHD (Observations=3,046 CHD events=159)		
	Odds ratio	95% confidence interval	P value	Odds ratio	95% confidence interval	P value	Hazard ratio	95% confidence interval	P value
C-reactive protein	1.05	0.99-1.12	0.105	1.07	0.96-1.20	0.223	1.06	0.95-1.18	0.313
Age	1.05	1.03-1.07	<0.001	1.02	0.99-1.06	0.163	1.16	1.03-1.31	0.017
Never smoked	1	-	-	1	-	-	1	-	-
Ex smoker	1.03	0.84-1.26	0.749	1.46	1.02-2.09	0.038	1.46	1.02-2.08	0.038
Current smoker	0.77	0.56-1.07	0.122	1.85	1.14-3.00	0.012	1.94	1.22-3.10	0.005
Systolic blood pressure[a]	0.78	0.71-0.86	<0.001	1.12	0.95-1.33	0.190	1.12	0.95-1.32	0.194
Total cholesterol	0.73	0.67-0.80	<0.001	0.98	0.85-1.13	0.796	0.99	0.86-1.14	0.922
Triglycerides (logged)	1.83	1.45-2.31	<0.001	1.47	0.99-2.19	0.056	1.45	0.99-2.13	0.060
Body mass index	1.03	1.01-2.31	0.009	1.00	0.97-1.04	0.988	1.00	0.97-1.04	0.922
FEV_1	0.55	0.45-0.68	<0.001	0.64	0.45-0.92	0.017	0.65	0.45-0.93	0.018
Diabetes	1.26	0.87-1.83	0.218	0.94	0.44-2.03	0.878	0.97	0.46-2.03	0.930
SES Class I	1	-	-	1	-	-	1	-	-
SES Class II	1.32	0.87-2.03	0.194	1.10	0.55-2.19	0.785	1.19	0.59-2.40	0.620
SES Class III (non-manual)	1.39	0.91-2.12	0.125	1.32	0.67-2.60	0.418	1.32	0.66-2.64	0.427
SES Class III (manual)	1.85	1.19-2.87	0.006	1.45	0.70-2.97	0.315	1.48	0.72-3.07	0.289
SES Class IV	1.89	1.23-2.90	0.004	1.05	0.51-2.16	0.903	1.16	0.56-2.40	0.692
SES Class V	2.04	1.10-3.80	0.024	1.00	0.31-3.26	0.999	1.05	0.33-3.38	0.932

Abbreviations: CHD, coronary heart disease; FEV_1, forced expiratory volume in one second; SES, socio-economic status.

[a] Parameter estimates are per standard deviation increase (25.1 mmHg)

Table 7.3: Basic descriptions of forced expiratory volume in one second, triglycerides and C-reactive protein from the datasets used for the prevalent CHD, incident CHD and time to incident CHD analyses.

Analysis	Observations	Events	Variable	Units of measurement	Mean	Standard deviation	Range
Prevalent CHD	3,620	574	FEV_1	Litres	1.98	0.52	0.45-5.30
			Triglycerides	Millimoles per litre	1.88	1.20	0.42-42.57
			Triglycerides (logged)	Log millimoles per litre	0.52	0.45	-0.87-3.75
			CRP	Milligrams per litre	3.47	5.84	0.16-112.00
			CRP (logged)	Log milligrams per litre	0.61	1.11	-1.81-4.72
Incident CHD and time to CHD	3,046	159	FEV_1	Litres	2.02	0.52	0.45-5.30
			Triglycerides	Millimoles per litre	1.85	1.24	0.42-42.57
			Triglycerides (logged)	Log millimoles per litre	0.50	0.45	-0.87-3.75
			CRP	Milligrams per litre	3.33	5.72	0.16-112.00
			CRP (logged)	Log milligrams per litre	0.57	1.11	-1.81-4.72

7.3.2. Repeatability measures for FEV_1, triglycerides and CRP

The values of the repeatability measures found in the literature for FEV_1, triglycerides and CRP are shown in Table 7.4. In order to estimate the long-term repeatability of FEV_1, triglycerides and CRP, all of the repeatability measures shown in Table 7.4 are derived from repeated measures that were not obtained on the same day. For FEV_1, titles and abstracts of 65 articles were reviewed. Two of these were considered relevant to this analysis, and a further two articles were obtained by reviewing their reference lists. For triglycerides, titles and abstracts of 283 articles were reviewed. Six of these were considered relevant to this analysis. No additional articles were found by reviewing the reference lists of these six articles. For CRP, titles and abstracts of 168 articles were reviewed. Five of these were considered relevant to this analysis. A further five results were obtained by reviewing the reference lists of these five articles.

Table 7.4: Repeatability measures of forced expiratory volume in one second, triglycerides and C-reactive protein, extracted from the literature, and the derived measurement error variances. In the British Women's Heart and Health Study, forced expiratory volume in one second is measured in litres, triglycerides is measured in millimoles per litre, and C-reactive protein is measured in log milligrams per litre.

Variable	Reference	Retest period	CV	ICC	Derived σ_e^2
FEV_1	Lebowitz *et al.*[191]	25-35 days	3.6		0.00509
	Rozas *et al.*[192]	5 days	2.8		0.00308
	Groth *et al.*[193]	15-180 days	4.7		0.00868
	Randell *et al.*[194]	1 week	2.8		0.00308
Triglycerides	Jacobs *et al.*[195]	1 week-1 year	~25		0.0247
	Godsland[196]	Weekly for 11-27 weeks	20.9		0.0171
	Bookstein *et al.*[197]	Monday, Wednesday and Friday of the same week	21.5		0.0181
	Brenner *et al.*[198]	1 week-1 year		0.78[a]	0.0455
	Marcovina *et al.*[199]	Fortnightly for 8 weeks	28.3		0.0320
	Egger *et al.*[200]	Average 1 year		0.46	0.0967
CRP	Clark *et al.*[201]	Fortnightly for 10 weeks	63.0		0.0721
	de Maat *et al.*[202]	Every 3 weeks for 6 months		0.86[b]	0.360
	de Maat *et al.*[203]	Every 4 weeks for 20 weeks	82[b]		0.526
	Schuit *et al.*[204]	1 year	32[b]		0.0801
	Macy *et al.*[205]	Every 3 weeks for 24 weeks	42.2		0.0311
	Sakkinen *et al.*[206]	Every 3 weeks for 24 weeks		0.77	0.125
	Riese[207]	5 days	18[b, c]		0.0253
	Ockene *et al.*[208]	Every 90 days for a year		0.78[b]	0.565
	Koenig *et al.*[209]	3 years		0.54[b]	1.18
	Broekmans[d]	6 months	37[b, c]		0.107

Abbreviations: CV, coefficient of variation; ICC, intra-class correlation coefficient; FEV1, forced expiratory volume in one second; CRP, C-reactive protein; σ_e^2*, measurement error variance.*

[a] *ICC for log triglycerides*

[b] *ICC or CV for log CRP*

[c] *Value obtained from Kluft et al.*[210]

[d] *No reference available. The dataset was obtained from Broekmans and analysed by Kluft et al.*[210]

The derived values of σ_e^2 for FEV_1 are all very small. This is due to the high repeatability of

FEV_1, indicated by low CVs. The method described in Section 7.2.3.1 is used to define the parameters of the normal distributions for the error variances. For FEV_1, the probability distribution chosen for the error variance is N(0.00588, 2.04×10^{-6}). The variance in this distribution is small due to the low variance between the repeatability measures values found in the literature. The probability of choosing negative error variances from this distribution is 1.9×10^{-5}. For triglycerides, the chosen probability distribution for the measurement error variance for triglycerides is N(0.0569, 4.12×10^{-4}), where the parameters are derived using the method in Section 7.2.3.1. The probability of choosing negative error variances from this distribution is 0.0026. The chosen probability distribution for the measurement error variance in CRP is N(0.603, 0.0868). The probability of choosing negative error variances from this distribution is 0.020. For all of these distributions, therefore, the coverage probabilities are not greatly affected by truncating the distribution at zero.

7.3.3. Crude analysis

First, results from the crude analyses are presented, where univariable models for the association between FEV_1, triglycerides, or CRP and three CHD outcomes (prevalent CHD at baseline, incident CHD, and time to incident CHD) are fitted. The naïve analyses give a crude OR for the effect of a unit increase in FEV_1 on prevalent CHD of 0.44 (0.37-0.53), for a unit increase in log triglycerides of 1.55 (1.28-1.87), and for a doubling of CRP of 1.17 (1.10-1.24). The crude OR for the effect of a unit increase in FEV_1 on incident CHD is 0.49 (0.36-0.67), for a unit increase in log triglycerides is 1.69 (1.21-2.35), and for a doubling of CRP is 1.16 (1.05-1.28). For time to incident CHD, the crude hazard ratio for the effect of a unit increase in FEV_1 on time to incident CHD is 0.52 (0.37-0.73), for a unit increase in log triglycerides is 1.66 (1.20-2.30), and for a doubling of CRP is 1.14 (1.03-1.26).

Table 7.5 shows the results of the crude analyses on prevalent CHD, incident CHD and time to CHD, allowing for measurement error in FEV_1, triglycerides or CRP. In the descriptions that follow, the width of the 95% CI or simulation interval is defined to be the difference between the logged values of the upper and lower limits for the intervals. This corresponds to the width of the 95% CI or simulation interval around the log OR or log hazard ratio.

For all results presented in this table, the 95% simulation interval from sensitivity analysis B is wider than the naïve 95% CI. This is a general result that does not apply only to the analyses considered in this section. Sensitivity analysis B incorporates both variation due to uncertainty over the true value of the measurement error variance, and sampling variation in the estimated parameter. As the variation due to uncertainty in the measurement error variance will always be greater than zero (for a sensitivity analysis), it follows that the 95% simulation interval will be wider than the 95% CI which includes only sampling variation.

The 95% simulation intervals from sensitivity analysis A are always narrower than the naïve 95% CI. This is not a general result, and may not apply to other analyses. It is possible for the variation due to uncertainty over the true value of the measurement error variance to be greater than the sampling variation. In this situation, the 95% simulation interval from sensitivity analysis A would be wider than the naïve 95% CI.

Table 7.5: Summary of effect estimates for a unit increase of FEV_1, a unit increase of log triglycerides, or a doubling of CRP on prevalent CHD, incident CHD or time to incident CHD from the cumulative probability distributions yielded by sensitivity analyses in which the amount of measurement error in FEV_1, triglycerides or CRP respectively is varied.

Outcome	Variable measured with error	Analysis	Median estimate	95% simulation interval
Prevalent CHD	FEV_1	Naïve analysis	0.44	(0.37-0.53)
		Sensitivity analysis A[a]	0.44	(0.43-0.44)
		Sensitivity analysis B[b]	0.44	(0.36-0.53)
	Triglycerides	Naïve analysis	1.55	(1.28-1.87)
		Sensitivity analysis A[a]	1.83	(1.61-2.27)
		Sensitivity analysis B[b]	1.83	(1.39-2.65[+])
	CRP	Naïve analysis	1.17	(1.10-1.24)
		Sensitivity analysis A[a]	1.23	(1.18-1.33)
		Sensitivity analysis B[b]	1.23	(1.13-1.39)
Incident CHD	FEV_1	Naïve analysis	0.49	(0.36-0.67)
		Sensitivity analysis A[a]	0.48	(0.47-0.49)
		Sensitivity analysis B[b]	0.48	(0.35-0.65)
	Triglycerides	Naïve analysis	1.69	(1.21-2.35)
		Sensitivity analysis A[a]	2.06	(1.78-2.69[+])
		Sensitivity analysis B[b]	2.06	(1.32-3.59[+])
	CRP	Naïve analysis	1.16	(1.05-1.28)
		Sensitivity analysis A[a]	1.22	(1.17-1.32)
		Sensitivity analysis B[b]	1.22	(1.06-1.44)
Time to incident CHD	FEV_1	Naïve analysis	0.52	(0.37-0.73)
		Sensitivity analysis A[a]	0.51	(0.51-0.51)
		Sensitivity analysis B[b]	0.51	(0.36-0.72)
	Triglycerides	Naïve analysis	1.66	(1.20-2.30)
		Sensitivity analysis A[a]	2.02	(1.74-2.60[+])
		Sensitivity analysis B[b]	2.02	(1.30-3.45[+])
	CRP	Naïve analysis	1.14	(1.03-1.26)
		Sensitivity analysis A[a]	1.19	(1.15-1.28)
		Sensitivity analysis B[b]	1.19	(1.04-1.41)

Abbreviations: FEV_1, forced expiratory volume in one second; CRP, C-reactive protein; CHD, coronary heart disease.

[a] Analysis allowing only for variability due to uncertainty about the true measurement error variance.

[b] Analysis allowing for both variability due to uncertainty about the true measurement error variance and sampling variation.

[+] Width of 95% confidence interval around the 97.5th percentile of the cumulative probability distribution>0.01.

For analyses estimating the association between FEV_1 and CHD, there is little variability in the cumulative probability distribution when only variation due to uncertainty over the true value of the measurement error variance is considered. This is shown by the small width of the 95% simulation interval for sensitivity analysis A. This is due to the small variance between the derived measurement error variances for FEV_1 from repeatability measures found in the

literature (see Table 7.4), and the correspondingly small variance in the probability distribution for the measurement error variance. This implies a small amount of uncertainty over the true value of this measurement error variance, and results in the narrow 95% simulation interval from sensitivity analysis A.

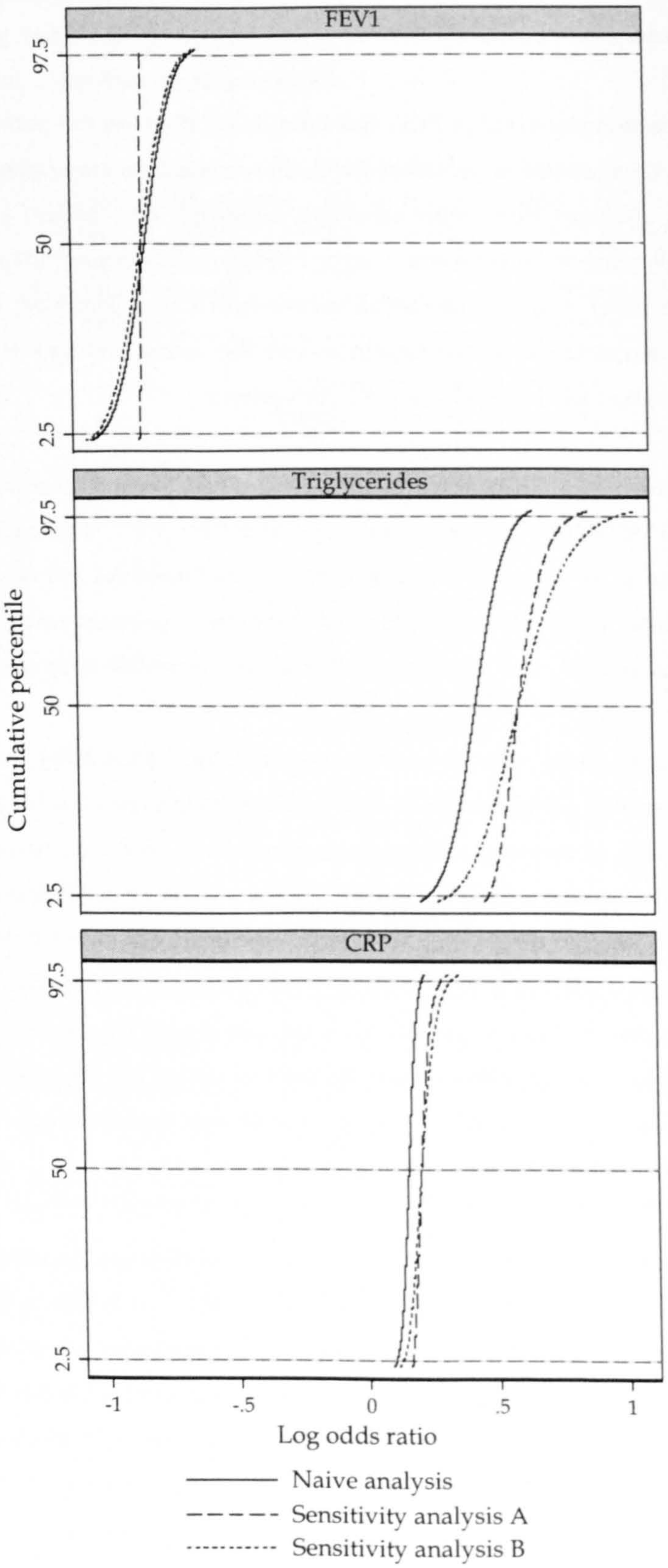
Allowing for measurement error in FEV_1 has little or no effect on the point estimates for all analyses. When the outcome is prevalent CHD, there is no difference between the naïve OR and the median estimates from either sensitivity analysis. For incident CHD and time to incident CHD, allowing for measurement error in FEV_1 has decreased the point estimate by 0.01. This small effect is due to the small measurement error variances derived from the repeatability measures found in the literature, and the correspondingly small mean in the probability distribution for the measurement error variance.

When allowing for measurement error in triglycerides or CRP, the sensitivity analyses have moved the point estimate away from the naïve estimated OR. Throughout this chapter, measurement error in all variables is assumed to be non-differential and random. Under this type of measurement error, the naïve estimates would be attenuated towards the null value. Allowing for measurement error therefore moves the point estimates away from the null value.

Figure 7.2 gives a graphical representation of the cumulative probability distribution for the association between FEV_1, triglycerides or CRP and prevalent CHD allowing for measurement error. A lateral shift of the sensitivity analysis results away from the naive analysis results corresponds to a change in the median estimate. The amount of lateral shift is due to the size of the mean in the probability distribution from which the error variance is drawn. A change in slope of the sensitivity analysis results compared with the naïve analysis changes the width of the 95% simulation interval for the log OR when compared with the naïve analysis. A steeper slope corresponds to a narrower 95% simulation interval for the log OR, and a shallower slope corresponds to a wider 95% simulation interval. The change in slope is due to the variance in the probability distribution from which the measurement error is drawn.

Considering the analysis of the effect of FEV_1 on prevalent CHD, there is very little lateral shift of the sensitivity analysis results from the naïve analysis. This can be seen in Table 7.5, as there is no change in the median estimate from the sensitivity analyses compared with the naïve point estimate. This is caused by the small amount of measurement error (small mean) in FEV_1. The cumulative probability distribution from sensitivity analysis A is an almost vertical line, which corresponds to the narrow 95% simulation interval seen in Table 7.5. The slope of the cumulative probability distribution for sensitivity analysis B is similar to that of the naïve analysis, which corresponds to the 95% simulation interval for sensitivity analysis B having a similar width to the 95% CI from the naïve analysis. The sensitivity analyses have shifted the

Figure 7.2: Graphical representation of sensitivity analysis results for the effect of a unit increase in FEV_1, a unit increase in log triglycerides, or a doubling of CRP on prevalent CHD, allowing for measurement error in FEV_1, triglycerides and CRP respectively.



cumulative probability distribution for the estimated association between triglycerides and prevalent CHD away from the naïve results, and away from the null value. This shift away from the null value is expected, as measurement error in triglycerides would attenuate the estimated association with CHD towards the null. Correspondingly, Table 7.5 shows an increase in the median estimates for both sensitivity analyses. The slope of the cumulative probability distribution of sensitivity analysis A is slightly steeper than the naïve analysis, which results in a narrower 95% simulation interval for the log OR. The slope of the cumulative probability distribution for sensitivity analysis B is shallower than that of the naïve analysis, and Table 7.5 shows a correspondingly wider 95% simulation interval compared with the width of the naïve 95% CI. The slopes of the cumulative probability distributions for the sensitivity analyses of triglycerides are shallower than those of FEV_1 because of the larger variance used in the probability distribution for the measurement error in triglycerides. The wide 95% CI around the upper limit of the 95% simulation interval for sensitivity analysis B, shown in Table 7.5, is due to the shallow slope of the cumulative probability distribution at this point. The patterns observed for the crude analysis of the effect of CRP on prevalent CHD are the same as those described for triglycerides.

The graphs for the other analyses in this section, and for all remaining analyses in this chapter, are not presented here, as the interpretation of the plots and relationship with the results presented in the tables are the same as described above. The graphs for the other analyses described in this section, and the analyses that are presented in Section 7.3.4, are available in Appendix 2.

7.3.4. Multivariable analysis

The multivariable sensitivity analysis for the effect of measurement error in FEV_1, triglycerides, CRP or a combination of these on the estimated association between CRP and the three CHD outcomes (prevalent CHD, incident CHD and time to incident CHD) will now be considered. The estimated associations from the naïve multivariable analyses, with no correction for error, give an OR for prevalent CHD of 1.05 (0.99-1.12) for a doubling of CRP, an OR for incident CHD of 1.07 (0.96-1.20) for a doubling of CRP, and a hazard ratio for time to incident CHD of 1.06 (0.95-1.18) for a doubling of CRP.

Table 7.6 summarises the sensitivity analysis results for prevalent CHD, incident CHD and time to incident CHD. As observed for the results of the crude analyses in Section 7.3.3, the 95% simulation intervals from sensitivity analysis A are all narrower than the corresponding naïve 95% CI, and the 95% simulation intervals from sensitivity analysis B are all at least as wide as the naïve 95% CI. When allowing for measurement error in FEV_1 only, the 95% simulation intervals for sensitivity analysis A are once again very narrow.

Allowing for measurement error in FEV_1 only has no effect on the point estimate for any of the analyses. This is once again due to the small derived measurement error variances from the repeatability measures found in the literature, and the correspondingly small mean in the probability distribution for the measurement error variance. The only differences observed between the naïve analysis results and the results of sensitivity analysis B occur when the outcomes are prevalent CHD or time to incident CHD, where the lower limit of the 95% simulation interval for sensitivity analysis B is slightly smaller than the lower limit of the naïve 95% CI.

For analyses with incident CHD or time to incident CHD as the outcome, allowing for measurement error in triglycerides only moves the point estimate slightly towards the null value. When the outcome is prevalent CHD, this effect is not observed, and allowing for measurement error in triglycerides has no effect on the point estimate.

Allowing for measurement error in triglycerides and FEV_1 generally produces results that are no different to the results obtained when allowing for measurement error in triglycerides alone. This is due to the lack of effect of allowing for measurement error in FEV_1 alone. The only difference between the results obtained when allowing for measurement error in triglycerides alone, and allowing for measurement error in triglycerides and FEV_1 occurs when the outcome is time to incident CHD. The upper limits for the 95% simulation intervals for sensitivity analysis A in these two cases are slightly different.

As expected, allowing for measurement error in CRP moves the point estimate away from the null value. Random and non-differential measurement error in the exposure will attenuate the naïve point estimate towards the null value, and therefore allowing for this error will move the point estimate away from the null value.

Allowing for measurement error in all three variables simultaneously moves the point estimate away from the null value, but not to the same extent as when allowing for measurement error in CRP alone. This demonstrates that the effects of measurement errors in the exposure and confounders can act in different directions, and that allowing for measurement error in the exposure variable only may produce too large a correction away from the null value. It is interesting to note that, although no attenuation of the point estimate is observed when the outcome is prevalent CHD and allowing for measurement error in triglycerides, the point estimate obtained when allowing for measurement error in all three variables is not as large as that observed when allowing for measurement error in CRP only. This suggests that allowing for measurement error in triglycerides does attenuate the effect estimate, but the attenuation effect is too small to be observed when allowing for measurement error in triglycerides only (or in triglycerides and FEV_1 together).

Table 7.6: Summary of estimated effects of doubling CRP on prevalent CHD, incident CHD or time to incident CHD from the cumulative probability distributions yielded by sensitivity analyses in which the amount of measurement error in FEV_1, triglycerides, CRP or a combination of these is varied.

Outcome	Variable measured with error	Analysis	Median estimate	95% simulation interval
Prevalent CHD	None	Naïve analysis	1.05	(0.99-1.12)
	FEV_1	Sensitivity analysis A[a]	1.05	(1.05-1.05)
		Sensitivity analysis B[b]	1.05	(0.98-1.12)
	Triglycerides	Sensitivity analysis A[a]	1.05	(1.04-1.06)
		Sensitivity analysis B[b]	1.05	(0.94-1.18)
	CRP	Sensitivity analysis A[a]	1.09	(1.06-1.15)
		Sensitivity analysis B[b]	1.09	(0.92-1.32)
	FEV_1 and triglycerides	Sensitivity analysis A[a]	1.05	(1.04-1.06)
		Sensitivity analysis B[b]	1.05	(0.94-1.18)
	CRP, FEV_1 and triglycerides	Sensitivity analysis A[a]	1.08	(1.05-1.13)
		Sensitivity analysis B[b]	1.08	(0.91-1.32)
Incident CHD	None	Naïve analysis	1.07	(0.96-1.20)
	FEV_1	Sensitivity analysis A[a]	1.07	(1.07-1.07)
		Sensitivity analysis B[b]	1.07	(0.96-1.20)
	Triglycerides	Sensitivity analysis A[a]	1.06	(1.04-1.07)
		Sensitivity analysis B[b]	1.06	(0.94-1.19)
	CRP	Sensitivity analysis A[a]	1.10	(1.07-1.17)
		Sensitivity analysis B[b]	1.10	(0.94-1.35)
	FEV_1 and triglycerides	Sensitivity analysis A[a]	1.06	(1.04-1.07)
		Sensitivity analysis B[b]	1.06	(0.94-1.19)
	CRP, FEV_1 and triglycerides	Sensitivity analysis A[a]	1.08	(1.05-1.15)
		Sensitivity analysis B[b]	1.08	(0.91-1.33)
Time to incident CHD	None	Naïve analysis	1.06	(0.95-1.18)
	FEV_1	Sensitivity analysis A[a]	1.06	(1.06-1.06)
		Sensitivity analysis B[b]	1.06	(0.94-1.18)
	Triglycerides	Sensitivity analysis A[a]	1.05	(1.03-1.06)
		Sensitivity analysis B[b]	1.05	(0.93-1.18)
	CRP	Sensitivity analysis A[a]	1.08	(1.06-1.14)
		Sensitivity analysis B[b]	1.08	(0.92-1.32)
	FEV_1 and triglycerides	Sensitivity analysis A[a]	1.05	(1.03-1.05)
		Sensitivity analysis B[b]	1.05	(0.93-1.18)
	CRP, FEV_1 and triglycerides	Sensitivity analysis A[a]	1.07	(1.04-1.11)
		Sensitivity analysis B[b]	1.07	(0.90-1.29)

Abbreviations: FEV_1, forced expiratory volume in one second; CRP, C-reactive protein; CHD, coronary heart disease.

[a] Analysis allowing only for variability due to uncertainty about the true measurement error variance.

[b] Analysis allowing for both variability due to uncertainty about the true measurement error variance and sampling variation.

7.4. Discussion

7.4.1. Summary of results

A method to allow for measurement error in continuous exposures and confounders has been developed, and the impact of measurement errors in two confounders and exposure on the estimated exposure-outcome association has been investigated. Although the effects of measurement error were small, it is preferable to investigate them rather than to ignore them

and report the results of a naïve analysis. The main findings in this chapter are that measurement error in FEV_1 is small and does not bias the estimated association between CRP and CHD, that measurement error in CRP attenuates the naïve estimate towards the null, but that the effect of residual confounding by triglycerides acts in the opposite direction and results in a smaller attenuation due to exposure measurement error.

7.4.2. Strengths and weaknesses

It was not possible to estimate error variances for FEV_1, triglycerides or CRP directly from the BWHHS data, as no repeated measures were made. External estimates, extracted from the literature from reported CVs or ICCs, were therefore used. These estimates are not necessarily transportable to the BWHHS data. This problem has been reduced by choosing an error variance from a probability distribution for which 95% of the choices will be in the range found in the literature, and not relying on a single external estimate. Nevertheless, the results presented in this chapter should not be interpreted as being perfectly corrected for measurement error, and should only be viewed as a sensitivity analysis of the possible effects of residual confounding by FEV_1 and triglycerides and measurement error in CRP.

Error variances for CRP, FEV_1 and triglycerides were drawn from normal distributions, based on repeatability values found in the literature. Other probability distributions are possible. Lash and Fink,[61] for example, drew values of sensitivity and specificity for their misclassified variable from triangular probability distributions. The choice of probability distribution is somewhat arbitrary. To investigate the impact of using different probability distributions, further analyses could be carried out in which the probability distribution is varied.

To derive the measurement error variances for the skewed variables (CRP and triglycerides), it was assumed that the skewed variable had a log-normal distribution, and that the log variable had a normal distribution. These assumptions are likely not to hold exactly, and there will therefore be some error in the derivation of the error variances for log triglycerides and CRP.

There are many assumptions required when a sensitivity analysis is undertaken. For example, in this chapter, classical and non-differential measurement error have been assumed, as well as assuming that the true measurement error variances for CRP, FEV_1 and triglycerides are taken from a normal distribution with parameters defined by the values found in the literature. While all assumptions may be validly criticised, this should not deter investigators from using sensitivity analysis methods. Considering measurement error, and attempting to quantify the effects through sensitivity analysis, is preferable to considering only sampling variation and reporting the results of a naïve analysis.

7.4.3. Implications

The results presented in this chapter have shown that allowing for measurement error in CRP consistently moves the median estimates from the sensitivity analyses away from the null value when compared with the naïve point estimates, and therefore that it is possible that the naïve estimates presented by previous studies into the association between CRP and CHD may be attenuated towards the null value. Residual confounding caused by errors in confounders, however, may act in the opposite direction, as was the case in this study with error in triglycerides. Residual confounding may therefore reduce any under-estimation of the exposure-outcome effect caused by measurement error in the exposure. In the case study presented in this chapter, accounting for residual confounding by triglycerides as well as measurement error in CRP resulted in a smaller change in the median estimate than was seen when accounting for the measurement error in CRP alone. As measurement errors in only two confounders were considered in this chapter, it is possible that considering measurement errors in other confounders could entirely cancel out the effects of exposure measurement error, or attenuate the naïve estimates towards the null. It therefore appears possible that the association between CRP and CHD may have been over-estimated in previous publications. As both under- and over-estimation of the association appears possible, the effects of measurement errors and misclassification of explanatory variables should be quantitatively assessed, as the combined effects of measurement errors in the exposure and confounders may be complex.

Allowing for measurement error in all variables included in an analysis is likely to be a time-consuming process. In this case-study, measurement error in FEV_1 had no effect on the estimated association between CRP and CHD. It would therefore appear unnecessary to allow for measurement error in all variables, but instead to restrict attention to the variables measured with moderate or large amounts of error.

Ideally, measurements should be free from error. This is often not possible in epidemiological studies. Internal or external validation studies, or repeated measurements should therefore be used to estimate measurement error variances, and these estimates used to correct the results of naïve analyses. The results from this chapter show that allowing for measurement error may have little or no effect on the naïve estimates. It is, however, better to consider measurement error in the exposure and confounders than to ignore it by using only a naïve analysis.

7.4.4. Future research

An extension to the analysis carried out in this chapter could include investigating the impact of measurement error or misclassification in all confounders simultaneously. This would require a further adaptation of the Lash and Fink sensitivity analysis method[61] to account for misclassification in polytomous confounders.

BLANK IN ORIGINAL

Part C.

Estimating treatment efficacy in the presence of departures from randomly allocated treatment in randomised controlled trials

BLANK IN ORIGINAL

Chapter 8.

Background literature: Methods for estimating treatment efficacy in the presence of departures from allocated treatment in randomised controlled trials

8.1. Introduction

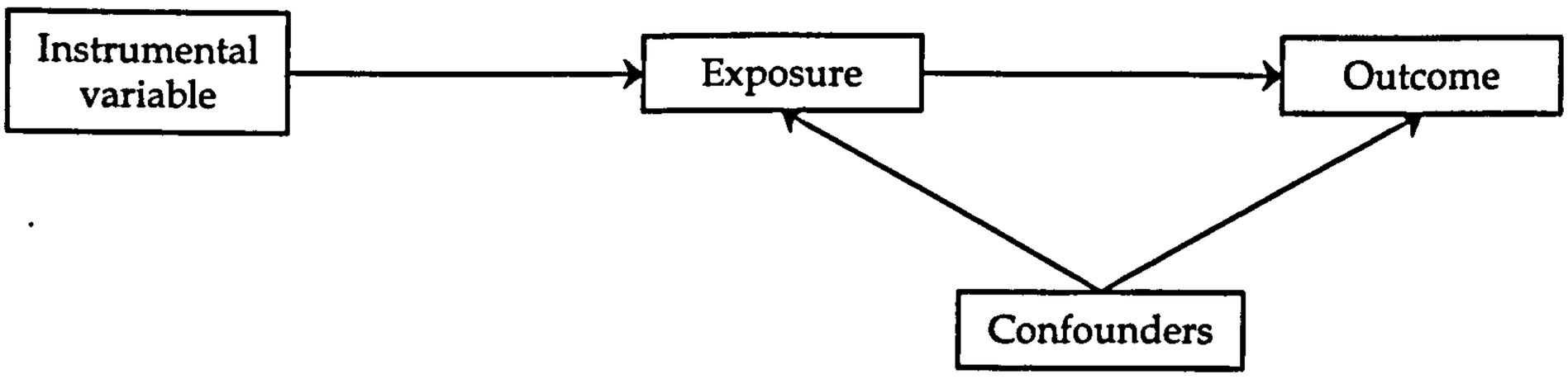
The primary method of analysis for randomised controlled trials (RCTs) is the intention-to-treat (ITT) analysis. Participants are analysed based on the treatment arm to which they were assigned, rather than treatment actually received. This method of analysis has the advantage that the groups compared should be similar with respect to both known and unknown prognostic factors providing sufficiently many participants are randomised, and therefore the effect estimate will not be biased by residual or unmeasured confounding. If there are departures from the allocated treatment the effect estimate is a mixture of the effect of treatment on compliers, and the lack of effect of treatment on non-compliers, and does not provide an estimate of treatment efficacy. Instead, it is an estimate of the effectiveness of treatment allocation. If the aim of the trial is to investigate the effect of treatment in a community intervention programme this effectiveness estimate may be the required parameter, assuming compliance rates in the community do not differ from those in the original trial. If the primary interest is instead on the efficacy of treatment, the ITT estimate will be biased when there are departures from the allocated treatment. If trial participants switch between treatment arms, but participants in the intervention arm receive more treatment than those in the control arm, the ITT analysis produces effect estimates that are biased towards the null. If trial participants switch to non-trial treatments, the ITT estimate for efficacy may be biased in either direction.

Alternative methods of analysis can be used in trials in which information on treatment actually received is recorded. An as-treated analysis analyses the effect of treatment received on outcome ignoring the randomisation. This is essentially an analysis of observational data and therefore suffers the same problems of residual and unmeasured confounding as observational studies. An alternative method of analysis is per-protocol analysis. In this method subjects are censored at the first time they depart from their randomised treatment. Per-protocol analysis therefore requires a binary definition of compliance, in order to specify the time at which the departure occurred. An underlying assumption of per-protocol analysis is that the two groups of subjects in each treatment arm that have not departed from their allocated treatment are comparable.[211] If this assumption does not hold, the per-protocol estimate will be subject to selection bias.

Statistical methods have been developed to estimate treatment efficacy in the presence of non-compliance while respecting the randomisation. These methods are all applications of instrumental variables methods, which are commonly used in economics. In epidemiological terminology, an instrumental variable is a variable that is independent of all confounders of the exposure-outcome association, is associated with exposure, but is independent of outcome given exposure and confounders.[212] This relationship is shown in Figure 8.1. In the context of RCTs, the randomisation is an instrumental variable, and can be used as such in analyses of the

effect of treatment actually received on outcome.

Figure 8.1: Instrumental variables in epidemiological studies.



A key concept in causal modelling is that of counterfactual (or potential) outcomes. Causal models aim to compare the outcome that would have been observed for a particular person under different treatment regimes. For example, the motivating example for this chapter is an analysis of the effect of sunscreen use on time to first basal cell carcinoma (BCC). The causal effect of sunscreen is obtained by comparing the time to BCC that would be observed if sunscreen was used every day with the time to BCC that would be observed if sunscreen was never used, with both outcomes relating to the same person. Only one outcome is observed for each person in a study, and all other outcomes are therefore contrary to fact and are termed counterfactual outcomes. This idea of defining causal effects by counterfactual outcomes has been called Rubin's Causal Model[213] following publications by Rubin.[214-217] The number of counterfactual outcomes for each individual in a study is potentially very large. This number is often limited by using the *stable unit-treatment value assumption* (SUTVA).[216] This states that the causal effect of treatment for an individual in the study is not affected by the treatment received by other individuals. In studies where treatment is binary (i.e. an individual either receives treatment or not) there are only two counterfactual outcomes for each person.

Background literature on methods which respect the randomisation and estimate treatment efficacy in the presence of non-compliance in RCTs are now described. Three specific methods are described in detail; estimation of complier average causal effects, rank preserving structural failure time models, and structural nested mean models.

8.2. Methods for estimating treatment efficacy in the presence of departures from allocated treatment in RCTs

8.2.1. Binary Compliance

8.2.1.1 Estimation of complier average causal effects

In many of the methods for analysing departures from allocated treatment in RCTs, it is assumed that compliance is all-or-nothing, i.e. that subjects either always take their allocated

treatment for the duration of the trial, or never take it. In these methods, participants are often divided into strata based on the treatment they would take if they were allocated to either the intervention or the control arm. *Compliers* are participants that would take the intervention treatment if and only if they were randomised to the intervention arm. *Always-takers* will always receive the intervention treatment, and *never-takers* will never receive the intervention treatment, regardless of the trial arm they are randomised to. *Defiers* will take treatment if and only if they are randomised not to receive it. These classifications have been called *compliance-types*[212] or *principal strata*.[218] Table 8.1 shows the principal strata in terms of randomised status and treatment received, and demonstrates that, without further assumptions, the principal strata are not observable. For example, people in the intervention arm who take the allocated treatment are either compliers or always-takers. If subjects in the intervention arm do not take the intervention treatment, they are either defiers or never-takers.

Table 8.1: Principal strata according to randomised status and treatment received.

Randomised group	Treatment received	
	Yes	No
Intervention	Always-taker Complier	Never-taker Defier
Control	Always-taker Defier	Never-taker Complier

Methods using the principal stratification framework often assume that there are no defiers (the *monotonicity* assumption). This allows identification of never-takers in the intervention arm and always-takers in the control arm (see Table 8.1). Using the fact that compliance-type is a pre-randomisation variable, and should therefore be independent of the randomised groups, the prevalence of never-takers (p_n) and the prevalence of always-takers (p_a) in the population can be estimated using the observed proportions in each trial arm. The prevalence of compliers (p_c) in the population can then be estimated as $p_c=1-p_n-p_a$.

Another common assumption is that if the randomisation does not affect the treatment actually received (as in the case of always and never-takers), then the randomisation does not affect the outcome. This is referred to as the *exclusion restriction* assumption. If interest lies in estimating an additive treatment effect on a continuous outcome, the exclusion restriction assumption states that the mean outcome for always-takers is the same in the intervention arm as in the control arm. The same argument applies to mean outcomes for never-takers. The only group, therefore, that has a different mean outcome in the intervention arm to the control arm is the group of compliers (assuming there are no defiers).

Estimation is usually of the efficacy of treatment among the compliers, and has been called the complier average causal effect (CACE). In the case of an additive treatment effect on mean

outcome, CACE estimation is intuitively simple. If the mean outcomes in the intervention and control groups are respectively denoted by $\overline{Y}_I$ and $\overline{Y}_C$, the ITT estimate in this situation equals $\overline{Y}_I - \overline{Y}_C$. In terms of mean outcomes in the principal strata, the ITT estimate equals $p_c(\mu_{c1} - \mu_{c0})$, where μ_{c1} and μ_{c0} are the mean outcomes among the compliers in the intervention and control groups respectively. There is no contribution to the ITT estimate from the always-takers and the never-takers because of the exclusion-restriction assumption of no difference in mean outcomes in the two randomised groups for always-takers and never-takers. The CACE estimate is therefore simply the ITT estimate divided by the estimate of p_c. A similar argument can be followed to obtain CACE estimates for binary or time-to-event outcomes and other measures of treatment efficacy, such as risk ratios.[211]

The description above shows an intuitively simple method to estimate the CACE point estimate. It is not clear, however, how to estimate confidence intervals (CIs) for the CACE estimate. A simple method, which ignores any uncertainty in estimation of additional parameters (such as p_c), is to follow the same method, but to use the ITT confidence limits where the ITT estimate would be used. Other possibilities are estimation via the delta method, using resampling techniques, such as the bootstrap or jackknife, or using Bayesian methods to obtain credible intervals from posterior distributions.

Some of the literature describing methods for obtaining CACE estimates is now described.

8.2.1.2 Background literature on CACE estimation

Sommer and Zeger[219] proposed a method to estimate the efficacy of vitamin A supplementation on mortality in a RCT with non-compliance. Children living in 225 randomly selected villages received vitamin A supplementation and were compared with children living in the remaining 225 villages that were not selected to receive supplementation. Their method was based on a comparison of compliant subjects in each arm of the trial. As there was no placebo control in the trial considered, compliance in the control arm was not observed. Four probabilities based on compliance type and outcome were therefore inferred for the control arm from data from the intervention arm; the probability of being a complier and alive, the probability of being a complier and dead, the probability for being a non-complier and alive, and the probability of being a non-complier and dead. To infer these probabilities, it was assumed that the expected rate of compliance was the same in both trial arms, and the expected mortality rate was the same in non-compliers in both arms. The number people in each of the four classifications in the control arm could then be derived from the probabilities. The estimate of efficacy was a ratio of the risk ratios among compliers in the intervention and control arms. As there was no placebo control, the estimated efficacy was a combination of the biologic efficacy of vitamin A and a placebo effect. The method applies only to all-or-nothing compliance and dichotomous outcome variables. Further assumptions are required to allow analysis of polytomous or

continuous compliance variables. The delta method was used to estimate the variance of the efficacy estimate. This method also does not allow for clustered responses, and therefore produces 95% CIs that are too precise. An extension to adjust variance estimates for clustering was proposed by Albert.[220]

Another method that estimates treatment efficacy among compliers where outcome is binary was described by Cuzick, Edwards and Segnan.[221] Their method incorporated non-compliance in both treatment arms, so that participants in the control arm were able receive the intervention treatment. The method was described in terms of the principal strata, although efficacy could only be estimated in the group of compliers with further assumptions such as equality of treatment effects in all principal strata, or the absence of defiers. Time to event outcomes may be incorporated into the method by dividing the follow-up period into several intervals and using a method similar to that described for binary outcomes. This method is a generalization of the method described by Sommer and Zeger,[219] and also used the delta method to estimate the variance of the efficacy estimate. Branson and Whitehead[222] developed a score test for the model[221] assuming all-or-nothing compliance and binary outcome data.

The principal stratification framework was also used by Angrist, Imbens and Rubin.[143] They described the application of instrumental variables to the problem of compliance in a randomised setting. They defined a Local Average Treatment Effect (LATE), which is the average causal effect for compliers, assuming that there are no defiers. The exclusion restriction and monotonicity assumptions were used to allow LATE estimation. Further assumptions, such as full compliance in the control arm, were required for the LATE estimate to be an estimate of the treatment effect in a subpopulation identifiable from the observed data. Standard errors for the IV estimates were obtained using a normal approximation to the sampling distribution of the ratio of the difference in estimated probability of outcome between the intervention and control groups, and the difference in probability of treatment received between the intervention and control groups. This model was extended by Baker[223] to account for survival outcomes, and used the delta method to estimate standard errors of efficacy estimates.

A Bayesian method of analysis which also used the principal stratification framework was described by Imbens and Rubin.[224] A major advantage of Bayesian methods in this setting is that the monotonicity assumption and exclusion restriction assumption are not required. The focus was on estimating the CACE, which is the same as the LATE estimate defined by Angrist, Imbens and Rubin.[143] The posterior distributions of the efficacy estimates were used to define the credible intervals. The method can be applied to both binary and continuous outcomes, and was illustrated using the vitamin A data of Sommer and Zeger.[219]

Hirano, Imbens, Rubin *et al.*[225] extended the Bayesian CACE estimation method[224] to allow for baseline covariates, and the exclusion restriction was divided into two components; an exclusion restriction assumption for always-takers, and one for never-takers. This contrasts with previous methods in which the exclusion restriction was applied to both always- and never-takers together. This allows a sensitivity analysis of the impact of violation of the exclusion restriction assumption. The plausibility of violations of the exclusion restriction in always-takers and never-takers should be assessed in the context of the analysis in question. Credible intervals were again derived from the posterior distributions of the estimates.

A further extension to the Bayesian CACE estimation methods[224, 225] was provided by Frangakis, Rubin and Zhou[218] to allow for clustered randomisation where non-compliance can occur at the individual level. This can occur in clustered encouragement designs, in which, for example, GP practices are randomised to either offer a treatment or not, and patients within each practise may or may not actually receive the treatment. The method was demonstrated with a binary outcome, and credible intervals for the efficacy estimates were obtained from the posterior distributions.

The issue of all-or-nothing compliance and non-ignorable missing outcomes in RCTs was considered by Frangakis and Rubin.[226] In this setting, ITT analysis is often based on the responders only and may therefore be biased. The authors proposed a method to obtain an ITT estimate that is unbiased in the presence of non-ignorable missing outcome data, based on a CACE estimate that can be obtained from the observed data. The delta method was used to estimate the variance of the efficacy estimate. The setting considered allows only for non-compliance in the intervention arm of the trial, although extensions to allow non-compliance in both trial arms are possible.

Levy, O'Malley and Normand[227] extended the method of Frangakis and Rubin[226] to allow adjustment for a continuous covariate in situations with all-or-nothing compliance. The model allows analysis of both continuous and binary outcomes, and CIs for the estimates were obtained using bootstrap resampling techniques. Simulation studies showed that the proposed estimators were unbiased when there were departures from the randomised treatment, with non-ignorable missing outcomes, and where there was a true effect of a continuous covariate on the outcome.

An all-or-nothing definition of compliance was used by Loeys and Goetghebeur[228] in the complier proportional hazards effect of treatment (C-PROPHET) model. This model allows for compliance in RCTs using a proportional hazards model, and non-compliance is assumed to only be possible in the intervention arm. Variance estimates for the estimated parameters are obtained using jackknife resampling techniques. Further work is required to include time-

Table 8.2: Summary of methods to estimate complier average causal effects.

Reference	Outcome			Clustered randomisation	Non-compliance in both arms	Adjustment for covariates	Missing outcomes	Bayesian approach	CI estimation
	Binary	Continuous	Survival						
Sommer *et al.*[219]	X								Delta method
Albert[220]	X			X					Delta method
Cuzick *et al.*[221]	X				X				Delta method
Angrist *et al.*[143]	X	X							Normal approximations
Baker[223]			X		X				Delta method
Imbens *et al.*[224]	X	X			X			X	Posterior distribution
Frangakis *et al.*[218]	X			X	X	X		X	Posterior distribution
Hirano *et al.*[225]	X				X	X		X	Posterior distribution
Frangakis *et al.*[226]		X	X				X		Delta method
Levy *et al.*[227]	X	X				X	X		Bootstrap resampling
Loeys *et al.*[228]			X						Jackknife resampling

Abbreviations: CI, confidence interval.

varying treatment, and partial compliance. A Stata command, written by Kim and White,[229] is available to implement this method.

8.2.1.3 Other methods

Ten Have, Joffe and Cary[230] proposed a method to obtain an estimate of the causal OR when there is treatment non-compliance. Rather than estimating a treatment effect among compliers, which is a common strategy in methods with binary compliance, the causal OR that would have been observed if everyone had received treatment was estimated. This approach does not require the monotonicity assumption, although it does require an assumption that treatment effects are the same in those who received treatment as those who did not. Simulation studies showed the parameter estimates were less biased than ITT or as-treated estimates, and the estimates had less bias than estimates obtained from a related method[231] (described later) when there was a strong relationship between compliance type and outcome.

The assumption of all-or-nothing compliance in principal stratification methods is often unrealistic. If it is possible for participants to switch treatment part way through a trial, and some participants do so, they cannot be classified into a principal stratum. Other analysis methods will be required under these circumstances.

8.2.2. Partial Compliance

Rather than assuming that compliance is all-or-nothing, methods have been proposed that will allow for a situation in which participants receive some, but possibly not all, of their allocated treatment.

Efron and Feldman[232] described a method for estimating a dose-response curve with partial compliance in a clinical trial of the effect of cholestyramine on cholesterol level. A model relating the outcome to the amount of treatment and the placebo response (the outcome that would be observed if the placebo were taken) was defined and used in the estimation of the dose-response curve. In general, the dose-response curve is not fully identified and therefore cannot be fully estimated, although a family of dose-response curves to which it belongs can be found. The curve can be fully estimated if either an assumption is made that there is no interaction between dose and placebo response, or if it is assumed that the dose-response curve is linear. Estimation of the dose-response curve also assumes that the placebo response and compliance are independent of the allocated treatment. In the example provided, however, compliance was better in the control group. This was accounted for by applying a simple adjustment to the compliance in the control group, and using this adjusted compliance when estimating the dose-response curve. The variance of the dose-response curve was estimated by a linear function of compliance separately in each treatment group. The coefficients for this linear function were obtained first by regressing outcome on compliance in the treatment and

control groups separately, and then regressing the squared residuals on compliance.

Zeger and Liang[233] extended Efron and Feldman's[232] method for cases in which compliance in the placebo and intervention groups is not comparable. In particular, they considered situations in which either the trial was not placebo controlled, or compliance with placebo was assumed to be irrelevant. They found that, under these circumstances, the dose-response curve could be estimated if a person's response to treatment does not depend on their tendency to comply, and if the compliance-response curve is linear. No consideration was given to variance estimation for the effect estimates.

The effect of errors in compliance measures in the model proposed by Zeger and Liang[233] was considered by Dunn.[234] Using simulation studies, measurement errors were shown to bias the required treatment-outcome effect estimates. A method for correcting for errors in compliance measures was proposed, using factor analysis. He concluded that measurement errors should be explicitly considered in compliance analyses, and that analysis models should be modified to account for such errors.

Data from the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT), previously analysed by Efron and Feldman,[232] were again used by Goetghebeur and Molenberghs.[235] In their method, compliance was a polytomous variable, with categories for less than 20%, between 20% and 60%, and greater than 60% compliance. Outcome was dichotomised, so that in the example a positive outcome corresponded to lowering cholesterol by at least 20 units. The dose-response curve was estimated in the different compliance subsets. It would be possible to extend the method to allow analysis of continuous outcomes with categorical compliance. The variance of the dose-response curve was estimated using solutions to the score equations.

The relationship between a continuous compliance variable and a continuous outcome variable was considered by Albert,[236] with a focus on determining a threshold dose beyond which treatment has no further effect. The dose response was defined as the difference between the outcome at a specific dose and zero dose, and was related linearly to dose less than a threshold value, with no further effect of dose above the threshold. Bootstrapping was used to estimate CIs for the threshold dose.

8.2.3. Repeated Compliance Measures

Another possible framework for data on compliance in RCTs is that there may be repeated measures over time. In this section, methods that allow repeated binary compliance measures are described.

8.2.3.1 Rank preserving structural failure time models

The rank preserving structural failure time model (RPSFTM) was originally proposed by Robins and Tsiatis.[237] Treatment received is related to failure time by an acceleration parameter that describes the amount lifetime is lengthened (or shortened) by treatment. A simple RPSFTM is

Equation 8.1: $$U_i = \int_0^{T_i} \exp(\beta X_i(t))dt$$

where U_i is the possibly counterfactual treatment-free failure time (i.e. the failure time that would be observed if no treatment was received), T_i is the observed lifetime, β is the acceleration parameter to be estimated, and $X_i(t)$ is treatment at time t. Using a binary variable to denote treatment received, the simple model assumes that the total time on (or off) treatment affects the causal treatment effect estimate, and therefore allows for repeated measures of treatment received. If $X_i(t)$ is a binary variable, Equation 8.1 reduces to

Equation 8.2: $$U_i = T_{0i} + e^{\beta} T_{1i},$$

where T_{0i} is the time that $X_i(t)$ equals 0, and T_{1i} is the time that $X_i(t)$ equals 1. The quantity e^{β} is known as the acceleration factor, and can be interpreted as the amount by which lifetime is accelerated by treatment. If $e^{\beta}<1$ (equivalently, $\beta<0$) treatment is protective and extends lifetime. If $e^{\beta}>1$ (equivalently, $\beta>0$) treatment is harmful and reduces lifetime, and if $e^{\beta}=1$ ($\beta=0$) there is no effect of treatment.

RPSFTMs respect the randomisation. If the ITT estimate has a 95% CI including the null value, the estimated parameter from the RPSFTM will also have a 95% CI including the null value. The RPSFTM parameter estimate will show an effect in the same direction as but further from the null than the ITT estimate.

The estimate of β from Equation 8.2, $\hat{\beta}$, is found by searching for the value of β for which the distributions of the treatment-free failure times, U, are equal between treatment arms. This process uses the randomisation as an instrumental variable to find the causal treatment effect estimate. The treatment-free failure times are calculated for various values of β using Equation 8.1 (or Equation 8.2 if the measures of treatment received are binary), and a log-rank test is used to test for the equality of the distributions between the randomised arms of the trial for each value of β. An interval bisection method or a grid search method can be used to find $\hat{\beta}$. For both methods, a range of values of β to search for the solution must be defined. The interval bisection method calculates the value of the log-rank statistic for the two values of β at either end of the search range. If the two log-rank statistics have opposite signs, there must be a solution between the two values of β. The interval is then bisected, and the process repeats until a value of β at which the log-rank statistic equals zero is found. If the initial two values of the log-rank statistic have the same sign the search range must be modified to find an interval which contains a solution. Alternatively, a grid search is possible. This evaluates the value of

the log-rank statistic at one end of the search range, and then increases the size of β in small steps until the other end of the search range is reached. The same methods can used to find the 95% CI for $\hat{\beta}$, using the points at which the log-rank statistic is 1.96 and -1.96. Although the grid search method is more computationally intensive than the interval bisection method, one major advantage of the grid search method is that multiple solutions can be identified. This occurs when the log-rank statistic is equal to, for example, 0 for different values of β. White, Babiker, Walker *et al.*[238] propose a method to deal with this problem. If the log-rank statistic crosses the required values (-1.96, 0 or 1.96) at the points $a_0, a_1, \ldots, a_n$, for even n, the estimate is taken to be $\sum_{i=0}^{n}(-1)^i a_i$. This corresponds to an average of the first and last crossing points, weighted by the total length of the interval between a_0 and a_n for which the log-rank statistic is positive and negative.

A further issue is that of censoring. Suppose, due to non-informative censoring such as administrative end of follow-up or random loss to follow-up, follow-up for subject i ends at time C_i. Then the censoring time for U_i, using Equation 8.1, is

$$D_i(\beta) = \int_0^{C_i} \exp(\beta X_i(t))dt.$$

Clearly, D_i is dependent on X_i, and therefore censoring of the treatment-free failure time is informative. For unbiased estimation, this dependence between D and X must be removed. This is achieved by recensoring U at the earliest censoring time possible over all treatment scenarios. If it is possible for X to always be 0 or always be 1, the recensoring time is given by

$$D_i^*(\beta) = \min(C_i, e^{\beta} C_i).$$

Using this recensoring time, events occurring close to the end of follow up will not be used. This loss of events can be viewed as follows. Participants with a particular treatment history, and an event occurring close to the end of follow-up, would not have had an event observed had their treatment history been more favourable. A different method, such as inverse-probability-of-censoring weighting, is required if censoring of the observed failure time is informative (see the appendix in Hernán, Cole, Margolick *et al.*[239]). Walker, White and Babiker[240] proposed a method to avoid recensoring, and therefore loss of information, when using a RPSFTM by specifying a bivariate frailty model for the association between treatment received and the latent failure time. As the proposed method is parametric, model misspecification may bias effect estimates.

RPSFTMs have also been described by White, Babiker, Walker *et al.*,[238] and White, Walker and Babiker[241] described a Stata command that allows implementation of the model. There was some work on bivariate models by White *et al.*,[238] although they concluded that the models proposed have little power or robustness for estimating more than one parameter. Further

examples of the use of RPSFTMs were provided by Mark and Robins,[242] and White, Walker, Babiker *et al.*[243]

While the acceleration factor has a simple interpretation, effect estimates for time to event analyses are more commonly presented as hazard ratios. White, Babiker, Walker *et al.*[238] proposed a method to obtain hazard ratios by correcting the event times using Equation 8.2, so that $T_i^* = T_{0i} + e^{\hat{\beta}} T_{1i}$ if the participant is randomised to the control arm, and $T_i^* = e^{-\hat{\beta}} T_{0i} + T_{1i}$ if the participant is randomised to the intervention arm. These corrected event times are the times that would have been observed had the subject remained on their randomly allocated treatment. The corrected event times are recensored in an analogous way to the recensoring described above, at the minimum censoring time of T^* over all possible treatment histories. For participants who do not change from their randomly allocated treatment, censoring of T^* therefore occurs at C_i. For participants who do change from their randomly allocated treatment, censoring of T^* occurs at $\min(C_i, C_i e^{\beta})$ in the control arm, and at $\min(C_i, C_i e^{-\beta})$ in the intervention arm. The corrected times are then used in a proportional hazards model to obtain hazard ratios.

This hazard ratio does not account for error in the acceleration parameter $\hat{\beta}$. White *et al.*[238] suggested that two 95% CIs should be presented for these hazard ratios. The first is a CI symmetric around the log hazard ratio, and with the same p-value as the ITT analysis. The second CI is the one obtained by taking 1,000 bootstrapped samples from each arm of the trial separately and estimating the hazard ratio in each bootstrap sample. It is possible that an estimate for the acceleration parameter will not be found in all bootstrapped datasets. In this situation, White *et al.*[238] calculated the corrected event time T^* using an extreme value of the acceleration parameter, and censored the resulting hazard ratio.

Problems may arise when estimating the causal parameter and 95% CI from a RPSFTM. For example, it may be impossible to find the 95% confidence limits, or there may be multiple possible values for the point estimate. These problems arise due to an inability of the model to distinguish true values of the causal parameter from other values. Mark and Robins[244] used simulation studies to investigate these problems. They found that the problems were a function of compliance, so that as compliance worsened the 95% CIs increased until the limits became infinite and the graph of the test statistic against the acceleration parameter became non-monotonic. If the probability of actually receiving the intervention treatment is equal in both randomised arms, the trial data cannot provide any information about the true effect of receiving treatment, as the survival curves in each arm will be equal in expectation.

The simulation studies presented by Mark and Robins[244] were extended by Korhonen, Laird and Palmgren,[245] with particular emphasis on situations in which there are unmeasured

confounders of the association between treatment received and outcome. They found that RPSFTMs can deal with unmeasured confounders at baseline, because the distribution of the confounders should be equal between the randomised arms.

RPSFTMs were extended to allow for clustering of outcomes in cluster randomised trials by Korhonen, Loeys, Goetghebeur *et al.*[246] and Loeys, Vansteelandt, and Goetghebeur.[247] In the first method, the treatment-free survival time was assumed to follow a proportional hazards model conditional on another covariate (age in their example), and then a robust covariance estimator was used to allow for the effect of clustering. The second method additionally allows for a cluster-specific frailty in the assumed proportional hazards model for treatment-free survival. The treatment effect was estimated as the value at which the treatment-free survival time is independent of the randomisation.

Loeys and Goetghebeur[248] extended the RPSFTM method to include baseline information. In the extension, a proportional hazards model related the baseline covariates to the treatment-free failure time. A RPSFTM was then used to estimate the effect of treatment on failure time. This method can lead to more precise estimates of the acceleration parameter than using log-rank tests, which is a common method of estimating the acceleration parameter in RPSFTMs.

RPSFTMs were extended by Matsui[249] to analyse time to repeated failure events in RCTs with non-compliance and informative censoring. An accelerated failure time model was defined for the time to each repeated outcome, and a separate model for the dependent censoring time. The method requires estimation of two parameters from the two accelerated failure time models. Vandebosch, Goetghebeur and Van Damme[250] also extended RPSFTMs to account for multiple failure events. They combined the RPSFTM proposed by Robins and Tsiatis[237] with the method proposed by Wei, Lin and Weissfeld[251] to deal with multiple failure events. Only non-informative censoring was considered.

8.2.3.2 Other methods

A model to analyse the effect of compliance in a trial with repeated binary outcomes was proposed by Sato.[252] In the method, exposure to the intervention treatment was binary and was measured on several different occasions. The focus was on estimating the causal risk difference, and estimation used the Mantel trend statistic. An estimator of the risk ratio was also presented, which is a generalisation of the estimator described by Cuzick, Edwards and Segnan.[221] The method calculates the predicted number of participants in the intervention group that would have had a certain number of events if they had taken the control treatment, for each possible compliance pattern and each possible number of events. In situations where there are more than three repeated measures of compliance the method therefore becomes difficult to use, as the number of possible compliance patterns increases exponentially. The

authors suggested only calculating the predicted numbers for the compliance patterns observed in the data.

A different method for analysing data with repeated compliance measures was proposed by Robins and Finkelstein.[253] In this method, a subject was regarded as censored the first time they stop their randomised treatment, switch treatments, or drop out of the study. Inverse-probability-of-censoring weights were then used to weight a log-rank test of the equality of survival distributions between the randomised arms. Repeated measures of compliance with the allocated treatment are therefore used in the model, but a binary definition of compliance is required to identify the first time a subject is non-compliant. This analysis method could result in loss of information, due to observed failures being artificially censored in those subjects that did not adhere to their randomised treatment. As loss of information increases, the widths of 95% CIs around the point estimates will also increase.

8.2.4. Repeated Measures and Partial Compliance

In the most general setting, trial data will contain repeated measures of partial compliance. Some methods have been developed to analyse such data.

8.2.4.1 Structural nested mean models

Robins[254, 255] proposed structural nested mean models (SNMMs) to correct for non-compliance in randomised trials. The model allows for mean treatment to be measured at several time points, and therefore uses repeated measures of partial compliance. The models were originally developed for continuous outcomes, which may be repeated, and can be applied in a fairly general RCT setting.

For continuous outcomes, the basic principle of SNMMs compares the observed outcome to the outcome that would have been observed if no treatment was received. The description of the method provided by Robins[254] is complex, but a more accessible description has been provided by Goetghebeur and Vansteelandt.[256] The aim is to estimate

Equation 8.3: $$E(Y - Y_0 | D, X, R = 1),$$

where Y is the observed outcome, Y_0 is the outcome that would have been observed if no treatment had been received, D is the treatment received, X is a vector of baseline covariates, and R is the randomisation indicator. Several assumptions are required to estimate Equation 8.3.

The first assumption is the *consistency assumption*. This states that, for both treatment arms and for given baseline covariates, the observed outcome when no treatment is received corresponds to Y_0, the possibly counterfactual treatment-free outcome.

The second assumption is the *randomisation assumption*. This states that Y_0 is independent of the

randomisation. This assumption is particularly useful when the control group cannot receive treatment, but is also useful when all trial participants can receive treatment. When subjects in the control arm cannot receive treatment the observed outcomes Y in the control arm will equal the treatment-free outcomes Y_0. The distribution of Y_0 can then be estimated in this arm and, using the randomisation assumption, the same distribution of counterfactual outcomes applies in the treatment arm.

Third, a causal model relating treatment received to the difference between Y and Y_0 must be assumed. This may be generally written as

$$E(Y - Y_0 | D, X, R) = DV_R'\varphi$$

where V_R is a vector function of the baseline covariates, and may be different in the different randomised arms, and φ is a vector of causal parameters to be estimated. Estimation is complex, as $Y-Y_0$ is only observed when no treatment is received and equals zero, and is achieved via the estimating equations:

$$\sum_{i=1}^{n} w(X_i)\{R_i - P(R_i = 1|X_i)\}\{Y_i - D_i V_R'(X_i)\varphi - q(X_i)\} = 0$$

where w(X) is a vector function and q(X) is a scalar function of baseline covariates. Robins[254] provided optimal choices for these functions. It is not clear how to estimate the variance of $\hat{\varphi}$. Goetghebeur and Lapp,[257] Lapp and Goetghebeur[258] and Loeys, Vansteelandt and Goetghebeur[247] have provided expressions for variance estimation in the specific settings considered in their papers.

The use of SNMMs in a more restricted setting of a placebo controlled trial was demonstrated by Goetghebeur and Lapp.[257] Fischer-Lapp and Goetghebeur[258] compared estimated parameters from SNMMs and ordinary least squares in situations in which the compliance measures were continuous. They found that the SNMM estimator was preferable, particularly where the treatment-outcome relationship suffered from unmeasured confounding.

The method proposed by Matsuyama,[259] which allows for repeated compliance and outcome measures, is a special case of a SNMM when outcome at time t is assumed to depend only on treatment at time t-1. Compliance and outcome were both binary variables, and estimation of both the causal risk difference and causal risk ratio were discussed.

SNMMs were extended by Loeys, Vansteelandt and Goetghebeur[247] to allow for different effects of treatment between clusters in cluster randomised trials. A cluster-specific parameter was included in the model which quantified how different the effect of treatment was in a specific cluster compared with the general population. The population treatment effect was estimated using a test of no difference in treatment effects between randomised arms. The variance of the

cluster-specific treatment parameter was estimated by testing for equal means of the square of the treatment-free outcomes between each randomised arm. These treatment-free outcomes may be counterfactuals, and were therefore estimated from the observed outcomes using the estimate of the population treatment effect.

Joffe and Bresinger[260] extended SNMMs to allow weights in analyses of the efficacy of treatment in RCTs. Throughout, outcome was a continuous variable, and compliance was a fixed proportion between one and zero. It was hypothesized that compliers provide more information about treatment efficacy, and therefore weighting by the compliance score (the effect of randomisation on treatment received) was used to increase the precision of efficacy estimates.

The impact of measurement error in treatment received when using linear SNMMs was investigated by Goetghebeur and Vansteelandt.[256] They found that parameter estimates were asymptotically unbiased in the presence of random treatment measurement error. Similarly, differential measurement error causes no bias in the SNMM parameter estimates. This contrasts with results from standard linear regression, in which random treatment measurement error will bias parameter estimates towards the null value, and differential measurement error may bias estimates in either direction (see Section 2.3). A classical measurement error model causes an increase in the variance of parameter estimates from SNMMs. Neither random nor systematic treatment measurement errors affect the α-level of hypothesis tests of no causal effect of treatment, but there may be a loss of power. If the treatment measurement error is systematic, with the bias known, a simple transformation is required in the treatment variable to enable unbiased estimation of the causal parameter of the SNMM. These results provide some reassurance that errors in compliance measures will not impact greatly on parameter estimates for analyses of this type.

8.2.4.2 Other methods

Participants in a RCT may often switch between the intervention and control arms, especially if the trial is placebo controlled. The method proposed by Nagelkerke, Fidler, Bernsen *et al.*[231] can account for such treatment crossover. They suggested that there is a variable, uncorrelated with randomised status, that when used in a regression analysis along with the variable describing treatment received removes all confounding that would arise from an as-treated analyses. This variable is equal to the residual of the regression of treatment received on randomised status. Their method may also apply to measures of partial compliance, if the treatment effect is assumed to be monotonically increasing with the amount of treatment received, and to time-varying treatments.

8.3. Summary

There are many methods available for analysing the effect of departures from allocated treatment in RCTs. Of all of the methods described in this chapter, CACE estimation, estimation using RPSFTMs and estimation using SNMMs appear to be the most well developed. The choice of model to use depends on the available compliance data and the outcome of interest. For example, in situations in which treatment is either received or not, such as for surgical interventions, CACE estimation methods can be used. These methods can be applied to a wide variety of outcomes. For more complex measures of compliance, for example repeated measures, or a single partial compliance measure, CACE estimation could still be used if the compliance data could be summarised as a single binary measure. In this situation, however, other methods to estimate treatment efficacy may be more appropriate. Rank preserving structural failure time models can be used with more complex compliance data where the total time using treatment can be calculated, but only allow for failure time outcomes. Structural nested mean models are also applicable to situations with complex compliance data, and may be used for continuous or binary outcomes.

Chapter 9.

Background literature: The association between sunlight and basal cell carcinoma

9.1. Introduction

The literature reviewed in this chapter is motivated by an analysis of the effect of sunscreen use on basal cell carcinoma (BCC), allowing for departures from allocated treatment in an RCT, that will be presented in Chapter 10. Basal cell carcinoma is the most common form of skin cancer, and is locally invasive but very rarely metastasizes.[261] The other two main types of skin cancer, squamous cell carcinoma (SCC) and melanoma, are not considered here.

In 2004, a review of environmental risk factors leading to the development of BCC was published.[262] Included among the numerous risk factors were ultraviolet (UV) radiation, smoking, occupational exposure to risk factors such as asphalt, mineral oils and organophosphate compounds, ingestion of arsenic containing medicines or arsenic contaminated water, and ionising radiation.

The risk factor of interest in this chapter is UV radiation or, more generally, exposure to sunlight. Evidence for the relationship between sunlight and BCC has been reviewed,[263] and BCCs were found to be more frequent in people residing closer to the equator, in those with sun-sensitive skin, in those with high levels of sun exposure, and in people with benign sun related skin lesions such as actinic keratoses. BCCs also occurred more frequently on sun-exposed sites, and were reduced by sun protection behaviour, such as wearing a hat. These six areas of evidence are reviewed below.

Papers were identified by searching Medline for articles under the medical subject headings of basal cell carcinoma and ultraviolet rays or sunlight. Titles and abstracts were reviewed and the relevant articles obtained. Additional articles were obtained by reviewing the reference lists of these papers.

9.2. The association between sunlight and BCC

9.2.1. Area of residence

There is some evidence that the risk of BCC increases as the distance of the region of residence from the equator decreases. For example, in a study conducted in the United States,[264] residents of California and Florida had an increased risk of BCC when compared with residents of the north eastern states of Massachusetts, Connecticut, New York, New Jersey, Pennsylvania and Maryland. The hazard ratio for BCC for residents of California was 1.51 (1.25-1.83), while that for residents of Florida was 2.03 (1.46-2.83). This evidence of an increased risk was not found when comparing residents of Texas with residents of the north eastern states, however, where the hazard ratio was 1.05 (0.75-1.47).

A cross-sectional prevalence study conducted in Australia investigated the effect of latitude of residence on prevalent BCC.[265] There was evidence that living closer to the equator increased the risk of BCC. Among men, the rate ratio for BCC when comparing those who lived at latitudes less than 29° south with those who lived at latitudes greater than 37° south was 3.2 (2.3-4.5). Among women, the rate ratio for BCC when comparing those who lived at latitudes less than 29° south with those who lived at latitudes greater than 37° south was 4.3 (2.9-6.5). Rate ratios and confidence intervals (CIs) comparing the intermediate category of between 29° south and 37° south with greater than 37° south were not provided.

A case-control study in Western Australia investigated the effect of ambient solar radiance at the place of residence on BCC.[266] There was some evidence that the risk of BCC increased as the accumulated global radiance over the whole year increased. Accumulated global radiance was measured in milliwatt hours per centimetre squared (mWh cm^{-2}). When compared with people with $8.8x10^{-5}$ mWh cm^{-2} or less, between $8.8x10^{-5}$ mWh cm^{-2} and $10.1x10^{-5}$ mWh cm^{-2} gave an odds ratio (OR) for BCC of 1.32 (0.69-2.55), between $10.1x10^{-5}$ mWh cm^{-2} and $11.4x10^{-5}$ mWh cm^{-2} gave an OR of 1.72 (0.72-4.09), and $11.4x10^{-4}$ mWh cm^{-2} or more gave an OR of 2.18 (0.82-5.82). The p-value for trend was 0.11. There was stronger evidence that the risk of BCC increased with accumulated global radiance over the warmer months. When compared with $5.8x10^{-5}$ mWh cm^{-2} or less, between $5.8x10^{-5}$ mWh cm^{-2} and $6.7x10^{-5}$ mWh cm^{-2} gave an OR of 1.40 (0.69-2.85), between $6.7x10^{-5}$ mWh cm^{-2} and $7.5x10^{-5}$ mWh cm^{-2} gave an OR of 2.45 (0.94-6.40), and $7.5x10^{-5}$ mWh cm^{-2} or more gave an OR of 3.44 (1.15-10.31). The p-value for trend was 0.02.

Analyses from the Health Professionals Follow-Up Study found evidence of an association between living in a region with high levels of sunlight and risk of BCC.[267] The hazard ratio for BCC was 1.48 (1.36-1.60) when compared with residents of areas with low levels of sunlight. There was little evidence that living in areas with medium levels of sunlight increased the risk of BCC when compared with areas with low levels of sunlight, as the hazard ratio for BCC was 1.01 (0.92-1.11). The p-value for trend, however, was less than 0.0001, providing strong evidence that the risk of BCC increases with increasing amounts of sunlight in the region of residence. There was also evidence that the risk of BCC was increased for people who had always lived in areas with high levels of sunlight (hazard ratio=1.46 (1.29-1.66)), and for people who only lived in an area with high levels of sunlight at the time of the study (hazard ratio=1.60 (1.40-1.83)) when compared with people who had always lived in areas with low levels of sunlight. There was little evidence that living in an area with high levels of sunlight as a child increased the risk of BCC, as the hazard ratio for BCC was 1.08 (0.80-1.45) when compared with people who had always lived in areas with low levels of sunlight. There was some evidence that the risk of BCC was increased for people who did not fit into any of the categories above (e.g. always living in areas with medium levels of sunlight) when compared with people who had always lived in areas with low levels of sun exposure, as the hazard ratio

Table 9.1: Summary of published associations between area of residence and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Hunter *et al.*[264]	73,366 women	**Region of residence**				
		North central states vs. north eastern states	ACDFG	0.92[a]	(0.75-1.13)	
		California vs. north eastern states	ACDFG	1.51[a]	(1.25-1.83)	
		Texas vs. north eastern states	ACDFG	1.05[a]	(0.75-1.47)	
		Florida vs. north eastern states	ACDFG	2.03[a]	(1.46-2.83)	
Marks *et al.*[265]	Men	**Latitude of residence**				
	(number unclear)	< 29° south vs. > 37° south	A	3.2	(2.3-4.5)	
	Women	**Latitude of residence**				
	(number unclear)	< 29° south vs. > 37° south	A	4.3	(2.9-6.5)	
Kricker *et al.*[266]	201 cases	**Accumulated global radiance during the year**				
	700 controls	**(mWh cm^{-2} $x10^{-5}$)**				
		8.8-10.1 vs. 0-8.8	AC	1.32	(0.69-2.55)	0.11
		10.1-11.4 vs. 0-8.8	AC	1.72	(0.72-4.09)	
		≥ 11.4 vs. 0-8.8	AC	2.18	(0.82-5.82)	
		Accumulated global radiance during the warmer months				
		(mWh cm^{-2} $x10^{-5}$)				
		5.8-6.7 vs. 0-5.8	AC	1.40	(0.69-2.85)	0.02
		6.7-7.5 vs. 0-5.8	AC	2.45	(0.94-6.40)	
		≥ 7.5 vs. 0-5.8	AC	3.44	(1.15-10.31)	
van Dam *et al.*[267]	317,084 person-years	**Current region of residence**				
	3,268 cases	Medium sun vs. low sun	ACDGI	1.01[a]	(0.92-1.11)	<0.0001
		High sun vs. low sun	ACDGI	1.48[a]	(1.36-1.60)	
	233,346 person-years	**Lifetime region of residence**				
	2,538 cases	Always high sun vs. always low sun	ACDGI	1.46[a]	(1.29-1.66)	
		Only currently high sun vs. always low sun	ACDGI	1.60[a]	(1.40-1.83)	
		High sun in childhood vs. always low sun	ACDGI	1.08[a]	(0.80-1.45)	
		Other vs. always low sun	ACDGI	1.09[a]	(0.99-1.20)	

A: Adjusted for general factors, such as age and sex, B: Adjusted for region of residence, C: Adjusted for skin sensitivity to sunlight, D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer, F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour, H: Adjusted for other environmental risk factors, I: Adjusted for ethnicity
[a] Hazard ratio

for BCC was 1.09 (0.99-1.20).

The results described above are summarised in Table 9.1, and appear to suggest that living in areas closer to the equator and with higher levels of ambient solar irradiance increases the risk of BCC. There are several hypotheses for this relationship. Kricker, Armstrong, English *et al.*[266] suggested that the association between area of residence and BCC could be due to increased intensity of exposure at areas of higher ambient solar irradiance, or that there is greater opportunity for outdoor activity in sunnier areas. Hunter, Colditz, Stampfer *et al.*[264] suggested that there might be increased exposure to ultraviolet-B (UVB) radiation in areas closer to the equator. The lack of evidence of an association between residence in Texas and BCC may be due to differences in sun exposure. Residents of Texas reported spending the least amount of time outside in summer of all of the states studied.[264] A further possible explanation for the apparent higher risk of BCC in sunnier areas is that the rates of detection of BCC in these areas are higher.[267]

9.2.2. Skin sensitivity to sun exposure

People with skin that tends to burn rather than tan have been shown to be at increased risk of BCC. For example, a study in Australia[268] found a rate ratio of 2.1 (p value=0.006) when comparing people who burned without tanning on their first sun exposure of the year with those who tanned only. There was, however, no control for confounding when estimating this association.

Evidence of an increased risk of BCC with skin that burns rather than tans was found in a cross-sectional study in the United States.[269] Comparing skin that burns with skin that tans, the OR for BCC was 2.72 (1.13-6.55).

In a case-control study in Western Australia, there was some evidence that the risk of BCC increased with increasing severity of skin reaction to sun exposure.[270] When compared with people who brown without burning, people who burn and then tan had an OR for BCC of 1.38 (0.81-2.34), people who have pain and peeling had an OR for BCC of 1.69 (0.98-2.91), and people who burn with blistering had an OR for BCC of 1.53 (0.79-2.99). The p-value for trend was 0.09.

A Canadian case-control study, however, found little evidence of an association between skin reaction to sun exposure and BCC.[271] Considering skin reaction to first sun exposure and when compared with people who never burn, people who burn after long exposure had an OR for BCC of 0.7 (0.4-1.3), people who burn after short exposure had an OR of 0.8 (0.4-1.5), and people who usually burn had an OR of 0.9 (0.4-1.9). There was also little evidence of an association between skin reaction to one week of sun exposure and BCC. When compared with people who tan without burning, people who tan when using sun protection had an OR of 1.1 (0.4-3.3),

people who burn then tan had an OR of 0.9 (0.6-1.5), and people who burn and never tan had an OR of 1.6 (0.8-3.2).

The Helios study in Southern Europe found evidence of an association between skin reaction to sun exposure and BCC.[272] The OR for BCC was 2.70 (2.10-3.47) when comparing people who burned without tanning with those who did not burn. Evidence of increased risk of BCC was also found among people who had a less severe skin reaction to sun exposure. Compared with people who did not burn, those who burned and then tanned had an OR for BCC of 1.49 (1.26-1.78).

The effect of skin reaction to two or more hours of sunlight as a child or adolescent was considered in a cohort study in the United States.[264] A hazard ratio for BCC of 2.41 (1.72-3.37) was found when comparing women who experienced painful sunburn with blisters with those who did not burn. Evidence of a harmful effect was also seen for less severe skin reactions to sun exposure as a child or adolescent. When compared with women who had no reaction to two hours or more of sun exposure, those who experienced painful burns had a hazard ratio for BCC of 1.70 (1.23-2.33), those who experienced burning had a hazard ratio of 1.87 (1.43-2.45), and those who experienced some redness had a hazard ratio of 1.40 (1.09-1.80). The p-value for trend was less than 0.001, providing strong evidence that the risk of BCC increases with increasing severity of skin reaction in childhood or adolescence.

Analysis of data from the Health Professionals Follow-Up Study found evidence that skin reaction to sun as an adolescent increased the risk of BCC.[267] When compared with people who tanned without burning in adolescence, those who burned than tanned had a hazard ratio for BCC of 1.51 (1.37-1.67), while those who burned painfully then peeled had a hazard ratio for BCC of 2.13 (1.90-2.38). The p-value for trend was less than 0.0001, providing strong evidence that the risk of BCC increases with increasing severity of skin reaction to sun exposure in adolescence.

An Italian case-control study also found evidence of an association between skin reaction to sun exposure and BCC.[273] The OR for BCC was 2.4 (1.7-3.8) for skin that burns frequently or always when first exposed to sun compared with skin that never burns. There was little evidence of an association between skin that burns occasionally when first exposed to sunlight, where the OR was 1.1 (0.8-1.4) when compared with skin that never burns. The p-value for trend was less than 0.001, providing strong evidence that the risk of BCC increases with increasing severity of skin sensitivity to sun exposure.

Strong evidence of a harmful association between skin sensitivity to sunlight and BCC was found in a Yugoslavian case-control study.[274] An OR of 4.34 (2.27-8.31) was found for skin that

Table 9.2: Summary of published associations between skin sensitivity to sun exposure and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Marks *et al.*[268]	1,981 subjects	**Skin reaction on first sun exposure of the year**				
		Burns only vs. tans only	A	2.1[a]	(p=0.006)	
Vitasa *et al.*[269]	808 men	**Skin reaction to sun exposure**				
		Burns vs. tans	ADEG	2.72	(1.13-6.55)	
Kricker *et al.*[270]	226 cases	**Propensity to burn**				
	1,018 controls	Burns then tans vs. tans without burning	AI	1.38	(0.81-2.34)	0.09
		Burns with pain and peeling vs. tans without burning	AI	1.69	(0.98-2.91)	
		Burns with blisters vs. tans without burning	AI	1.53	(0.79-2.99)	
Gallagher *et al.*[271]	226 cases	**Skin reaction to first sun exposure**				
	406 controls	Burns after long exposure vs. never burns	AG	0.7	(0.4-1.3)	
		Burns after short exposure vs. never burns	AG	0.8	(0.4-1.5)	
		Usually burns vs. never burns	AG	0.9	(0.4-1.9)	
	226 cases	**Skin reaction to 1 week of exposure**				
	405 controls	Tans with protection vs. tans without burning	AG	1.1	(0.4-3.3)	
		Burns then tans vs. tans without burning	AG	0.9	(0.6-1.5)	
		Burns and never tans vs. tans without burning	AG	1.6	(0.8-3.2)	
Hunter *et al.*[264]	73,366 women	**Skin reaction to ≥2 hours sunlight as a child or adolescent**				
		Some redness vs. no burn	ABCDFG	1.40[b]	(1.09-1.80)	<0.001
		Burn vs. no burn	ABCDFG	1.87[b]	(1.43-2.45)	
		Painful burn vs. no burn	ABCDFG	1.70[b]	(1.23-2.33)	
		Painful burn with blisters vs. no burn	ABCDFG	2.41[b]	(1.72-3.37)	
Suárez-Varela *et al.*[275]	Men (number unclear)	**Skin reaction to sun exposure**				
		Burns without tanning vs. tans without burning	A	1.4[c]	(1.0-2.2)	
	Women (number unclear)	**Skin reaction to sun exposure**				
		Burns without tanning vs. tans without burning	A	0.7	(0.4-1.3)	
Zanetti *et al.*[272]	1,541 cases	**Skin reaction to sun exposure**				
	1,790 controls	Burns then tans vs. tans without burning	ABCDG	1.49	(1.26-1.78)	
		Burns without tanning vs. tans without burning	ABCDG	2.70	(2.10-3.47)	

Table 9.2 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
van Dam *et al.*[267]	249,366 person-years	**Skin reaction to sun exposure as an adolescent**				
	2,713 cases	Burns then tans vs. tans without burning	ABCDEGI	1.51[b]	(1.37-1.67)	<0.0001
		Painfully burns then peels vs. tans without burning	ABCDEGI	2.13[b]	(1.90-2.38)	
Naldi *et al.*[273]	528 cases	**Skin reaction to sun exposure**				
	509 controls	Burns occasionally vs. never burns	AG	1.1	(0.8-1.4)	<0.001
		Burns frequently/always vs. never burns	AG	2.4	(1.7-3.8)	
Vlajinac *et al.*[274]	200 cases	**Skin reaction to sun exposure**				
	399 controls	Usually burns with little or no tan	CDEGH	4.34	(2.27-8.31)	

A: Adjusted for general factors, such as age and sex
B: Adjusted for region of residence
C: Adjusted for skin sensitivity to sunlight
D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer
F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour
H: Adjusted for other environmental risk factors
I: Adjusted for ethnicity
[a] Rate ratio
[b] Hazard ratio
[c] Includes all non-melanoma skin cancers as outcome

usually burns with little or no tanning, although the reference group for this analysis was unclear.

Additional evidence for an association between skin sensitivity to sun exposure and BCC comes from a Spanish case-control study.[275] The OR for non-melanoma skin cancer (NMSC) (i.e. SCC and BCC) was 1.4 (1.0-2.2) when comparing men with a tendency to burn without tanning with those with a tendency to tan. Little evidence of an association between skin sensitivity and NMSC in women was found, however, as the OR comparing women who tend to burn without tanning with those who tan without burning was 0.7 (0.4-1.3). As analyses included SCCs the association between skin sensitivity to sun exposure and BCC alone is not clear.

The results described above are summarised in Table 9.2. The evidence does not indicate a clear association between skin sensitivity to sun exposure and BCC, although it does seem to suggest that, if there is an association between skin sensitivity to sun exposure and BCC, that it is not protective. One possibility for the mixed results is that residual and unmeasured confounding is a problem in many of the estimated associations.

9.2.3. Location of skin cancer

The most common areas of the body for BCC to occur are the areas usually exposed to sunlight, such as the head, and least common are rarely exposed areas such as the abdomen.

In a cross-sectional prevalence study in Australia,[276] 92% of prevalent BCCs occurred on the head and neck. When this population was followed up for incident cases of NMSC 94% of incident BCCs occurred on the head and neck, with 6% occurring on the hands and forearms.[268] Only BCCs occurring on the head, neck, forearms and hands were considered in these studies.

A further cross-sectional prevalence study was conducted in Australia in 1990.[265] BCCs were most common on the head and neck, accounting for 66% of prevalent BCCs among men and 69% of prevalent BCCs among women. The next most common site was the trunk, with 21% and 14% of prevalent BCCs in men and women respectively. The upper limbs accounted for 8% of the BCCs occurring in men and 9% of the BCCs occurring in women, while the lower limbs accounted for 4% and 6% in men and women respectively. The location of the BCC was not determined for 1% of cases in men and 2% of cases in women.

In a cross-sectional prevalence survey of 808 white male fishermen residing in Maryland in the United States, 78.3% of prevalent BCCs occurred on the face, head or neck.[269] The next most common location was the trunk with 13.3% of prevalent BCCs, followed by the upper extremities with 8.4% of prevalent BCCs. No BCCs were found on the lower extremities.

Basal cell carcinomas were shown to be most common on the head in a case-control study in Southern Europe,[272] accounting for 78.1% and 76.9% of cases in men and women respectively. The next most common site was the trunk, with 14.1% of cases in men and 10.1% of cases in women. BCCs on the neck, abdomen, upper limbs and lower limbs were less common, with each site accounting for less than 3% of cases in men, and 4.2% of cases in women.

In a case-control study in Spain,[275] NMSC was most common on the head, with 94.2% of the cases occurring there. The trunk was the next most common location and accounted for 2.9% of the cases. The neck, upper extremities and lower extremities accounted for 1.1%, 1.1% and 0.7% of the cases respectively. These results include both SCC and BCC.

In a case-control study in Italy,[273] 63.3% of cases had a BCC on the head or neck. The trunk was the next most common site with 27.7% of cases. Fewer BCCs were found on the limbs (7.2%), or at multiple sites (1.9%).

Similar results were reported for a case-control study in Southern Germany,[277] where 58.2% of the BCCs in the cases were on the head. Acral BCCs (i.e. BCCs occurring on peripheral parts of the body, such as the ears or fingers) were the next most common, comprising 31.0% of all BCCs. BCCs on the trunk, lower extremities and upper extremities were much less common, comprising 6.6%, 2.8% and 1.4% of all cases respectively.

These results are summarised in Table 9.3. The hypothesis that exposure to sunlight is a major cause of BCC is supported by the fact that, in every study described, the majority of BCCs occurred on the head and neck, which are commonly exposed to sunlight. That a higher proportion of BCCs occur on the trunk, an area less commonly exposed to sunlight, appears to contradict this hypothesis. Several reasons have been advanced for this. Marks *et al.*[265] suggested that an increase in the proportion of BCCs observed on the trunk and upper limbs is due to a change in fashions, and less clothing being worn on these areas. Squamous cell carcinomas do not display the same distribution on the body as BCC. Green *et al.*[278] suggested that the difference may be due to the fact that a lower amount of exposure to UV radiation is required for BCCs to develop than for SCCs. The pattern of sun exposure may also be important. Intermittency of exposure has been proposed as an explanation for the high proportions of BCCs seen on the trunk, an area that is infrequently exposed to sunlight.[279, 280] This idea was extended by Rosso *et al.*,[281] who suggested that risk of BCC may be affected by how directly UV radiation reaches the basal layers. In intermittently exposed sites, such as the trunk, the skin protection against sunlight does not develop as quickly, and therefore the UV radiation enters the basal layer more directly. Once skin protection has developed, SCCs tend to form instead of BCCs. Of course, there may be a simple dose-response relationship between BCC and sun exposure in which increasing exposure increases the number of BCCs that

Table 9.3: Summary of literature describing the location of basal cell carcinomas.

Reference	Number	Head and neck	Trunk	Abdomen	Upper limbs	Lower limbs	Limbs	Multiple sites	Acral	Unknown
Marks *et al.*[276]	38 cases	92%			8%					
Marks *et al.*[268]	118 cases	94%			6%					
Marks *et al.*[265]	438 cases	67%	19%		8%	5%				1%
Vitasa *et al.*[269]	60 cases	78.3%	13.3%		8.4%	0.0%				
Zanetti *et al.*[272]	1,549 cases	79.4%	12.4%	2.7%	1.6%	2.7%				
Suárez-Varela *et al.*[275]	276 cases	95.3%	2.9%[a]		1.1%[a]	0.7%[a]				
Naldi *et al.*[273]	528 cases	63.3%	27.7%				7.2%	1.9%		
Walther *et al.*[277]	213 cases	58.2%	6.6%		1.4%	2.8%			31.0%	

[a] *Includes all non-melanoma skin cancers*

develop. The lack of BCCs on the lower limbs, including the hands, may then be due to the skin on the hands providing more protection against BCC.[280]

9.2.4. Level of sun exposure

9.2.4.1 Sunburn

Sunburn can be considered a marker of sun exposure, and evidence has been found for a harmful association between sunburn and BCC. In a cohort study in the United States,[264] evidence of an increased risk of BCC was found when comparing six or more painful sunburns on the face and arms with none, with a hazard ratio of 1.90 (1.50-2.40). There was also evidence of increased hazard ratios for smaller numbers of lifetime sunburns. Compared with no severe or painful sunburns on the face and arms, women with three to five sunburns had a hazard ratio of 1.34 (1.05-1.71), and those with one or two sunburns had a hazard ratio of 1.18 (0.94-1.48). The p-value for trend was less than 0.001, providing strong evidence that the risk of BCC increases with increasing numbers of lifetime sunburns.

There was, however, little evidence of an association between painful sunburn and BCC in a cohort study in Queensland, Australia.[282] When compared with people with no painful sunburns, people with one had an OR for BCC of 0.5 (0.2-1.4), people with between two and five painful sunburns had and OR for BCC of 0.6 (0.3-1.5), and people with six or more painful sunburns had an OR for BCC of 1.0 (0.4-2.5).

There was similarly little evidence of an association between lifetime sunburn pain and BCC in a Canadian case-control study.[271] When comparing people who had ever experienced two days or more of sunburn pain over their lifetime with people who had never experience such sunburn pain, the OR for BCC was 0.9 (0.6-1.3). There was some evidence that sunburn pain over the last decade had a harmful association with BCC. Comparing people who had experienced two days or more of sunburn pain over the last decade with people who had not, the OR for BCC was 2.3 (0.8-6.6).

A case-control study in Western Australia[279] found evidence of an increased risk of BCC when comparing people with between three and ten sunburns to people with none with an OR of 1.75 (1.08-2.85). There was also some evidence of an association when people with eleven or more sunburns were compared with those with none, with an OR for BCC of 1.50 (0.99-2.26). There was little evidence of an association between having one or two sunburns and BCC, where the OR was 1.09 (0.67-1.79). There was some evidence of a harmful association between blistering sunburns and BCC. When comparing people with one or two blistering sunburns to people with no sunburn, the OR for BCC was 1.60 (0.92-2.79). There was, however, little evidence of an association when comparing painful sunburn with no blisters with no sunburn (OR = 1.05 (0.71-

Table 9.4: Summary of published associations between sunburn and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Hunter *et al.*[264]	73,366 women	**Sunburn to face and arms**				
		1-2 vs. none	ABCDFG	1.18[a]	(0.94-1.48)	<0.001
		3-5 vs. none	ABCDFG	1.34[a]	(1.05-1.71)	
		6+ vs. none	ABCDFG	1.90[a]	(1.50-2.40)	
Green *et al.*[282]	1,770 subjects	**Number of painful sunburns**				
		1 vs. none	AEG	0.5	(0.1-1.4)	
		2-5 vs. none	AEG	0.6	(0.3-1.5)	
		≥ 6 vs. none	AEG	1.0	(0.4-2.5)	
Gallagher *et al.*[271]	226 cases	**≥ 2 days of sunburn pain in the lifetime**				
	406 controls	Ever vs. never	AGI	0.9	(0.6-1.3)	
		≥ 2 days sunburn pain the last decade				
		Yes vs. no	AGI	2.3	(0.8-6.6)	
Kricker *et al.*[279]	192 cases	**Frequency of painful sunburn**				
	700 controls	1-2 times vs. none	AC	1.09	(0.67-1.79)	
		3-10 times vs. none	AC	1.75	(1.08-2.85)	
		≥ 11 times vs. none	AC	1.50	(0.99-2.26)	
		Frequency of blistering sunburn				
		Sunburn with no blisters vs. no sunburn	AC	1.05	(0.71-1.55)	
		1-2 blistering burns vs. no sunburn	AC	1.60	(0.92-2.79)	
		≥ 3 times vs. no sunburn	AC	1.24	(0.69-2.24)	
Zanetti *et al.*[272]	1,549 cases	**Lifetime number of sunburns**				
	1,795 controls	One vs. none	ABCG	1.13	(0.94-1.36)	0.031
		Two vs. none	ABCG	1.30	(0.92-1.84)	
		3+ vs. none	ABCG	1.30	(0.95-1.78)	
van Dam *et al.*[267]	250,530 person-years	**Lifetime number of blistering sunburns**				
	2,723 cases	1-2 vs. none	ACGI	1.14[a]	(1.00-1.30)	<0.0001
		3-5 vs. none	ACGI	1.20[a]	(1.05-1.38)	
		6-9 vs. none	ACGI	1.33[a]	(1.14-1.54)	
		≥ 10 vs. none	ACGI	1.49[a]	(1.30-1.71)	

Table 9.4 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Walther *et al.*[277]	146 cases 264 controls	**Sunburn 20 years before diagnosis** Ever vs. never	ABDEG	3.6	(1.9-6.8)	

A: Adjusted for general factors, such as age and sex
B: Adjusted for region of residence
C: Adjusted for skin sensitivity to sunlight
D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer
F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour
H: Adjusted for other environmental risk factors
I: Adjusted for ethnicity
* *Hazard ratio*

1.55)), or when comparing three or more blistering sunburns with no sunburn (OR = 1.24 (0.69-2.24)).

In a case-control study in Southern Europe, some evidence was found of an association between number of lifetime sunburns and BCC.[272] Compared with people with no sunburns, people with one sunburn had an OR for BCC of 1.13 (0.94-1.36), those with two sunburns had an OR of 1.30 (0.92-1.84), and those with three or more sunburns had an OR of 1.30 (0.95-1.78). The p-value for trend was 0.031, providing some evidence that the risk of BCC increases with an increasing number of lifetime sunburns.

Analyses from the Health Professionals Follow-Up Study found evidence of an association between lifetime number of blistering sunburns and risk of BCC.[267] When compared with people with no blistering sunburns in their lifetime, people with one or two had a hazard ratio for BCC of 1.14 (1.00-1.30), people with between three and five had a hazard ratio for BCC of 1.20 (1.05-1.38), people with between six and nine had a hazard ratio for BCC of 1.33 (1.14-1.54), and people with ten or more had a hazard ratio for BCC of 1.49 (1.30-1.71). The p-value for trend was less than 0.0001, providing strong evidence that the risk of BCC increases with increasing numbers of lifetime blistering sunburns.

Evidence of an association between sunburn 20 years before the diagnosis of BCC and BCC was found in a German case-control study.[277] The OR for BCC was 3.6 (1.9-6.8) when comparing people who experienced sunburn with those who did not.

The estimated associations between sunburn and BCC described above are summarised in Table 9.4.

9.2.4.2 Occupational sun exposure

There is some evidence that occupational sun exposure is associated with BCC. A cohort study in Australia[268] found a rate ratio for BCC of 1.6 (p value=0.030) when comparing people who work outdoors with those who work indoors. No control for confounding was made when estimating this association.

A cohort study in Queensland, Australia,[282] found some evidence of an association between occupational exposure to sunlight and BCC when comparing people who work indoors and outdoors to people who work mainly indoors (OR=1.5 (0.9-2.9)). There was little evidence of an association between working mainly outdoors and BCC, where the OR for BCC was 1.3 (0.6-2.8) when compared with people who work mainly indoors.

Similarly, little evidence of an association between occupational sun exposure and BCC was

found in a Canadian case-control study.[271] When compared with less than 3.5 hours per week of occupational sun exposure in summer, between 3.5 and 13.9 hours per week of exposure had an OR for BCC of 1.0 (0.6-1.8), between 14 and 24.9 hours per week of exposure had an OR for BCC of 1.3 (0.8-2.3), and 25 hours per week or more had an OR for BCC of 1.4 (0.8-2.4). The p-value for trend was greater than 0.05.

There was little evidence of an association between occupational exposure to sunlight and BCC in a case-control study in Western Australia.[266] When compared with 14,700 hours of accumulated sun exposure on working days since the age of 15, between 14,800 hours and 27,700 hours had an OR for BCC of 1.25 (0.79-1.97), between 27,800 hours and 49,300 hours had an OR for BCC of 1.17 (0.72-1.90), and 49,400 hours or more had an OR for BCC of 0.86 (0.50-1.51). The p-value for trend was 0.46.

Little evidence of an association between occupational sun exposure and BCC was found in a European case-control study.[281] When compared with people with less than 7,200 hours of sun exposure during outdoor work, between 7,200 hours and 54,720 hours of exposure gave an OR for BCC of 1.02 (0.84-1.24), and more than 54,720 hours gave an OR for BCC of 1.00 (0.78-1.30).

Some evidence of a harmful association between occupational sun exposure and BCC was found by an Italian case-control study.[273] When comparing intermediate levels of occupational sun exposure with low levels, the OR for BCC was 1.4 (0.9-2.0). There was, however, little evidence of an association when comparing high levels of occupational sun exposure with low levels, where the OR for BCC was 0.9 (0.6-1.3). There was little evidence of an increasing risk of BCC with increasing levels of occupational sun exposure, as the p-value for trend was 0.69.

Evidence of an association between outdoor work in summer and BCC was found in a Yugoslavian case-control study,[274] with an OR for BCC of 3.95 (1.62-9.66). The reference category for this analysis was unclear.

A German case-control study also found evidence of a harmful association between occupational sun exposure and BCC.[277] The OR for BCC was 2.4 (1.3-4.7) when comparing frequent or occasional occupational UV exposure with rare or no UV exposure.

Evidence of an association between occupational sun exposure and NMSC was found among men in a Spanish case-control study.[275] The OR for BCC was 5.3 (3.1-9.2) when comparing chronic occupational sun exposure of more than 4.8 hours per day with less than 1.7 hours per day. Comparing men with between 3.6 and 4.8 hours per day of occupational sun exposure with those with less than 1.7 hours per day, the OR for NMSC was 2.5 (1.5-4.3). There was little evidence of an association between occupational sun exposure and NMSC when comparing

Table 9.5: Summary of published associations between occupational sun exposure and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Marks *et al.*[268]	1,981 subjects	**Occupational exposure**				
		Outdoors vs. indoors occupation	A	1.6[a]	(p=0.030)	
Green *et al.*[282]	1,770 subjects	**Occupational exposure**				
		Indoors and outdoors vs. mainly indoors	AEG	1.5	(0.8-2.9)	
		Mainly outdoors vs. mainly indoors	AEG	1.3	(0.6-2.8)	
Gallagher *et al.*[271]	226 cases	**Lifetime mean occupational sun exposure**				
	406 controls	**(hours/week in summer)**				
		3.5-13.9 vs. < 3.5	AGI	1.0	(0.6-1.8)	> 0.05
		14.0-24.9 vs. < 3.5	AGI	1.3	(0.8-2.3)	
		≥ 25 vs. < 3.5	AGI	1.4	(0.8-2.4)	
Kricker *et al.*[266]	201 cases	**Sun exposure on working days since age 15**				
	700 controls	**(hours)**				
		14,800-27,700 vs. 0-14,700	AC	1.25	(0.79-1.97)	0.46
		27,800-49,300 vs. 0-14,700	AC	1.17	(0.72-1.90)	
		≥ 49,400 vs. 0-14,700	AC	0.86	(0.50-1.51)	
Rosso *et al.*[281]	1,549 cases	**Sun exposure during outdoor work**				
	1,795 controls	**(hours)**				
		7,200-54,720 vs. < 7,200	ABCDG	1.02	(0.84-1.24)	
		> 54,720 vs. < 7,200	ABCDG	1.00	(0.78-1.30)	
Naldi *et al.*[273]	526 cases	**Occupational sun exposure**				
	511 controls	Intermediate vs. low	AG	1.4	(0.9-2.0)	0.69
		High vs. low	AG	0.9	(0.6-1.3)	
Vlajinac *et al.*[274]	200 cases	**Occupational sun exposure**				
	399 controls	Outdoor work during summer	CDEGH	3.95	(1.62-9.66)	
Walther *et al.*[277]	146 cases	**Occupational exposure**				
	264 controls	Frequent/sometimes vs. rare/never	ABEG	2.4	(1.3-4.7)	

Table 9.5 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Suárez-Varela *et al.*[275]	Men (number unclear)	**Occupational exposure (hours/day)**				
		1.7-3.6 vs. <1.7	A	1.2[b]	(0.6-2.2)	
		3.6-4.8 vs. < 1.7	A	2.5[b]	(1.5-4.3)	
		> 4.8 vs. < 1.7	A	5.3[b]	(3.1-9.2)	
Suárez-Varela *et al.*[275]	Women (number unclear)	**Occupational exposure (hours/day)**				
		1.7-3.6 vs. < 1.7	A	2.1[b]	(0.8-5.6)	
		3.6-4.8 vs. < 1.7	A	0.8[b]	(0.2-4.2)	
		> 4.8 vs. < 1.7	A	0.8[b]	(0.1-8.2)	

A: Adjusted for general factors, such as age and sex
B: Adjusted for region of residence
C: Adjusted for skin sensitivity to sunlight
D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer
F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour
H: Adjusted for other environmental risk factors
I: Adjusted for ethnicity
[a] *Rate ratio*
[b] *Includes all non-melanoma skin cancers*

men with exposure of between 1.7 and 3.6 hours per day with men with less than 1.7 hours per day of exposure (OR=1.2 (0.6-2.2)). Little evidence of an association between occupational sun exposure and NMSC was found for women, however, for any of the categories of occupational sun exposure. When compared with less than 1.7 hours per day of chronic occupational exposure, women with between 1.7 and 3.6 hours per day of exposure had an OR for BCC of 2.1 (0.8-5.6), women with between 3.6 and 4.8 hours per day of exposure had an OR for BCC of 0.8 (0.2-4.2), and women with more than 4.8 hours per day of exposure had an OR for BCC of 0.8 (0.1-8.2). These results include both SCCs and BCCs and therefore the effect of sun exposure on BCCs alone cannot be established.

The estimated associations between occupational sun exposure and BCC described above are summarised in Table 9.5.

9.2.4.3 Recreational sun exposure

Some evidence has been found that recreational sun exposure, such as at weekends or while on holiday, is also associated with BCC. However, little evidence of an association between sun exposure during leisure was found in a cohort study in Queensland, Australia.[282] When compared with people who spend their leisure time mainly indoors, people who spend their leisure time indoors and outdoors had an OR for BCC of 1.0 (0.4-2.2), and people who spend their leisure time mainly outdoors had an OR for BCC of 0.6 (0.3-1.3).

In a case-control study in Western Australia,[279] there was evidence that the risk of BCC was increased for all categories of lifetime hours of sun exposure while on holiday when compared with less than 602 hours of exposure. Comparing people with between 602 and 2,268 hours of sun exposure with those with less than 602 hours of exposure, the OR for BCC was 1.65 (1.01-2.70). Comparing people with between 2,268 and 3,794 hours of sun exposure to those with less than 602 hours of exposure gave an OR for BCC of 1.68 (1.00-2.80). The greatest risk occurred in the group with greater than 3,794 hours of sun exposure, with an OR of 1.85 (1.09-3.13) when compared with people with less than 602 hours of sun exposure on holidays. The p-value for trend was 0.04, providing some evidence that the risk of BCC increases with increasing number of hours of sunlight exposure while on holiday. There was some evidence of an association between sunbathing and BCC. Comparing people who had sunbathed between one and 200 times during their lifetime with people who had never sunbathed, the OR for BCC was 1.57 (0.98-2.51). There was, however, little evidence of an association between more frequent sunbathing and BCC. Comparing people with between 201 and 700 occasions of sunbathing in their lifetime with those with none, the OR for BCC was 1.08 (0.68-1.72), while for people with between 700 and 9,000 lifetime instances of sunbathing the OR for BCC was 1.02 (0.63-1.64).

Intermittency of sun exposure was also investigated in this study.[279] Intermittency was

measured as the proportion of days exposed to sunlight on non-working days compared with the total number of days exposed. There was little evidence of an association between intermittency of sun exposure and BCC among people between 20 and 24 years old. When compared with intermittency of 20% or less, between 21% and 44% intermittency gave an OR for BCC of 1.35 (0.83-2.18), between 45% and 89% intermittency gave an OR for BCC of 1.06 (0.59-1.90), and between 90% and 100% intermittency gave an OR for BCC of 1.71 (0.88-3.34). The p-value for trend was 0.2. There was also little evidence of an association between intermittency and BCC in people between 25 and 39 years old. Compared with intermittency of 19% or less, between 20% and 39% intermittency gave an OR for BCC of 1.31 (0.82-2.11), between 40% and 89% intermittency gave an OR for BCC of 1.40 (0.83-2.35), and between 90% and 100% intermittency gave an OR for BCC of 1.13 (0.57-2.22). The p-value for trend was 0.45. There was weak evidence of an association between intermittency of sun exposure in the 10 years prior to diagnosis and BCC. When compared with 24% intermittency or less, between 25% and 49% intermittency had an OR for BCC of 1.75 (1.15-2.66), between 50% and 99% intermittency had an OR for BCC of 2.10 (1.25-3.54), and 100% intermittency had an OR for BCC of 1.22 (0.65-2.31). The p-value for trend was 0.10. The evidence of an association between intermittency between 11 and 30 years prior to diagnosis and BCC was also weak. When compared with 29% or less intermittency, between 30% and 56% intermittency had an OR for BCC of 1.58 (0.96-2.61), between 57% and 99% intermittency had an OR for BCC of 1.42 (0.77-2.64), and 100% intermittency had an OR for BCC of 1.75 (0.89-3.45). The p-value for trend was 0.14.

Data from the same study showed evidence of an increased risk of BCC with increased hours of sun exposure on non-working days since the age of 15.[266] An OR for BCC of 1.74 (1.03-2.95) was found when comparing more than 16,300 hours of exposure with less than 6,700 hours. When compared with less then 6,700 hours of exposure, between 11,200 hours and 16,200 hours of exposure had an OR of 1.76 (1.07-2.90), and between 6,700 hours and 11,100 hours of exposure had an OR of 1.72 (1.05-2.79). The p-value for trend was 0.04, providing some evidence that the risk of BCC increases with increasing hours of sun exposure on non-working days since the age of 15.

The Helios study also found evidence of a harmful association between recreational sun exposure and BCC.[281] Comparing people with more than 3,398 hours of lifetime sun exposure during holidays with people with no holiday sun exposure produced an OR for BCC of 1.47 (1.18-1.83). When compared with people with no holiday sun exposure, people with between 1,324 hours and 3,398 hours of exposure had an OR for BCC of 1.10 (0.88-1.39), people with between 280 hours and 1,323 hours of exposure had an OR for BCC of 1.26 (1.01-1.56), and people with less than 280 hours of exposure had an OR of 1.20 (0.97-1.48). The p-value for trend was 0.036, providing evidence that the risk of BCC increases with increasing sun exposure on

holidays.

Sun exposure at the beach on holidays showed some evidence of a harmful association with BCC. When compared with people with no sun exposure at the beach, less than 2,464 hours of exposure had an OR for BCC of 1.12 (0.95-1.32), and more than 2,464 hours of exposure had an OR for BCC of 1.47 (1.18-1.84).

There was strong evidence that the risk of BCC increased with increasing time doing water sports. An OR for BCC of 1.47 (1.04-2.07) was found when comparing more than 2,112 hours of sun exposure during water sports with no exposure. When compared with no sun exposure during water sports, less than 2,112 hours of exposure gave an OR for BCC of 1.45 (1.18-1.79).

For sun exposure during outdoor sports in general, the OR for BCC when comparing less than 288 hours of exposure with none was 1.22 (0.99-1.51). There was, however, little evidence of an association with BCC for higher levels of sun exposure during outdoor sports. When compared with no sun exposure during outdoor sports, between 288 and 1,008 hours of exposure had an OR for BCC of 1.10 (0.89-1.51), between 1,009 and 3,420 hours of exposure had an OR for BCC of 1.07 (0.86-1.32), and more than 3,420 hours of exposure had an OR for BCC of 1.01 (0.84-1.28). The p-value for trend was 0.552, providing little evidence that the risk of BCC increases with increasing sun exposure during outdoor sports.

There was also little evidence of an association between sun exposure during sports that take place in the mountains or in air (i.e. skiing, climbing, hiking, flying, hang-gliding and parachuting) and BCC. Compared with people with no sun exposure during sports in the mountains or air, people with less than 140 hours of exposure had an OR for BCC of 1.22 (0.85-1.77), people with between 140 and 504 hours of exposure had an OR for BCC of 1.14 (0.79-1.66), people with between 505 and 1,887 hours of exposure had an OR for BCC of 1.06 (0.72-1.54), and people with more than 1,887 hours of exposure had an OR for BCC of 1.04 (0.72-1.52). The p-value for trend was 0.438, providing little evidence that the risk of BCC increases with increasing amounts of sun exposure during sports in the mountains or air.

A Yugoslavian case-control study[274] found evidence of a harmful association between an average of seven or more weeks per year of seaside holidays and BCC, with an OR of 1.81 (1.24-2.64). The reference category for this analysis was unclear.

In an Italian case-control study[273] intermediate amounts of lifetime recreational sun exposure gave an OR for BCC of 1.8 (1.3-2.6), and high lifetime recreational sun exposure gave an OR of 1.6 (1.1-2.3), when compared with low lifetime levels of sun exposure. The p-value for trend was less than 0.001, providing strong evidence that the risk of BCC increases with increasing lifetime recreational sun exposure.

Some additional evidence of a harmful effect of recreational sun exposure on BCC is provided by a Spanish case-control study of NMSC.[275] In men, high levels of sun exposure (>0.12 hours per day) during open air activities had an OR for BCC of 1.7 (1.1-2.8) when compared with low levels of exposure (<0.005 hours per day), although there was little evidence of an association between sun exposure during open air activities and NMSC for an intermediate amount of sun exposure (0.005-0.12 hours per day) (OR=0.7 (0.5-1.3)). An OR of 2.1 (1.2-3.9) was found when comparing more than 0.4 hours per day of sun exposure during holidays with less than 0.16 hours per day. Again there was little evidence of an association between sun exposure and NMSC for lower levels of sun exposure. Compared with less than 0.16 hours per day of sun exposure during holidays, between 0.16 and 0.26 hours per day of sun exposure had an OR for BCC of 0.8 (0.5-1.6), and between 0.26 and 0.40 hours per day of exposure had an OR for BCC of 1.3 (0.7-2.6).

There was some evidence of an association between high levels of sun exposure during outdoor activities and NMSC among women, as the OR for BCC was 2.1 (0.9-4.7) when comparing more than 0.12 hours per day of exposure with less than 0.005 hours per day of exposure. There was, however, little evidence of an association between intermediate levels of sun exposure during outdoor activities (0.005-0.12 hours per day), with an OR for BCC of 1.4 (0.5-3.8) when compared with less than 0.005 hours per day of exposure. There was also little evidence of an association between sun exposure during holidays and NMSC for women. When compared with less than 0.16 hours per day of exposure, between 0.16 and 0.26 hours per day of exposure had an OR for BCC of 1.7 (0.7-4.3), between 0.26 and 0.40 hours per day of exposure had an OR for BCC of 1.2 (0.6-2.5), and more than 0.40 hours per day of exposure had an OR for BCC of 1.4 (0.6-3.2). These analyses included both SCCs and BCCs, and therefore the effect of recreational sun exposure on BCCs alone cannot be established.

In contrast to the results described above, a case-control study of Canadian men found evidence of a protective association between recreational sun exposure and BCC.[271] Comparing a lifetime mean of 8.5 hours per week or more of recreational sun exposure in summer with a mean of less than 2.8 hours per week gave an OR for BCC of 0.4 (0.2-1.0). When compared with less than 2.8 hours per week of recreational sun exposure in summer, between 2.8 and 5.5 hours per week of exposure had an OR for BCC of 0.9 (0.5-1.7), and between 5.6 and 8.4 hours per week of exposure had an OR for BCC of 0.6 (0.3-1.3). The p-value for trend was 0.03, providing evidence that the risk of BCC decreases with decreasing amounts of lifetime recreational sun exposure.

The estimated associations between recreational sun exposure and BCC described above are summarised in Table 9.6.

Table 9.6: Summary of published associations between recreational sun exposure and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Green *et al.*[282]	1,770 subjects	**Leisure exposure**				
		Indoors and outdoors vs. mainly indoors	AEG	1.0	(0.4-2.2)	
		Mainly outdoors vs. mainly indoors	AEG	0.6	(0.3-1.3)	
Kricker *et al.*[279]	192 cases 700 controls	**Lifetime sun exposure on holiday (hours)**				
		602-2,268 vs. 0-602	AC	1.65	(1.01-2.70)	0.04
		2,268-3,794 vs. 0-602	AC	1.68	(1.00-2.80)	
		> 3,794 vs. 0-602	AC	1.85	(1.09-3.13)	
		Lifetime frequency of sunbathing				
		1-200 times vs. none	AC	1.57	(0.98-2.51)	
		201-700 times vs. none	AC	1.08	(0.68-1.72)	
		701-9,000 times vs. none	AC	1.02	(0.63-1.64)	
	201 cases 700 controls	**Intermittency of sun exposure (ages 20-24)**				
		21-44% vs. 0-20%	ACD	1.35	(0.83-2.18)	0.2
		45-89% vs. 0-20%	ACD	1.06	(0.59-1.90)	
		90-100% vs. 0-20%	ACD	1.71	(0.88-3.34)	
		Intermittency of sun exposure (ages 25-39)				
		20-39% vs. 0-19%	ACD	1.31	(0.82-2.11)	0.45
		40-89% vs. 0-19%	ACD	1.40	(0.83-2.35)	
		90-100% vs. 0-19%	ACD	1.13	(0.57-2.22)	
		Intermittency of sun exposure 11-30 years prior to diagnosis				
		30-56% vs. 0-29%	ACD	1.58	(0.96-2.61)	0.14
		57-99% vs. 0-29%	ACD	1.42	(0.77-2.64)	
		100% vs. 0-29%	ACD	1.75	(0.89-3.45)	

Table 9.6 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Kricker *et al.*[279]	201 cases 700 controls	**Intermittency of sun exposure 10 years prior to diagnosis**				
		25-49% vs. 0-24%	ACD	1.75	(1.15-2.66)	0.10
		50-99% vs. 0-24%	ACD	2.10	(1.25-3.54)	
		100% vs. 0-24%	ACD	1.22	(0.65-2.31)	
Kricker *et al.*[266]	201 cases 700 controls	**Sun exposure on non-working days since age 15 (hours)**				
		6,700-11,100 vs. < 6,700	AC	1.72	(1.05-2.79)	0.04
		11,200-16,200 vs. < 6,700	AC	1.76	(1.07-2.90)	
		> 16,300 vs. < 6,700	AC	1.74	(1.03-2.95)	
Rosso *et al.*[281]	1,549 cases 1,795 controls	**Lifetime sun exposure during holidays (hours)**				
		<280 vs. none	ABCG	1.20	(0.97-1.48)	0.036
		280-1,323 vs. none	ABCG	1.26	(1.01-1.56)	
		1,324-3,398 vs. none	ABCG	1.10	(0.88-1.39)	
		> 3,398 vs. none	ABCG	1.47	(1.18-1.83)	
		Lifetime sun exposure at the beach on holidays (hours)				
		< 2,464 vs. none	ABCDG	1.12	(0.95-1.32)	
		≥ 2,464 vs. none	ABCDG	1.47	(1.18-1.84)	
		Lifetime sun exposure during water sports (hours)				
		< 2,112 vs. none	ABCDG	1.45	(1.18-1.79)	
		≥ 2,112 hours vs. none	ABCDG	1.47	(1.04-2.07)	
		Lifetime sun exposure during outdoor sports (hours)				
		< 288 vs. none	ABCG	1.22	(0.99-1.51)	0.552
		288-1,008 vs. none	ABCG	1.10	(0.89-1.51)	
		1,009-3,420 vs. none	ABCG	1.07	(0.86-1.32)	
		> 3,420 vs. none	ABCG	1.01	(0.84-1.28)	

Table 9.6 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Rosso *et al.*[281]	1,549 cases 1,795 controls	**Lifetime sun exposure during sports in the air or mountains (hours)**				
		<140 vs. none	ABCG	1.22	(0.85-1.77)	0.438
		140-504 vs. none	ABCG	1.14	(0.79-1.66)	
		505-1,887 vs. none	ABCG	1.06	(0.72-1.54)	
		> 1,887 vs. none	ABCG	1.04	(0.72-1.52)	
Vlajinac *et al.*[274]	200 cases 399 controls	**Recreational sun exposure**				
		Average of 7+ weeks per year of seaside holidays	CDEGH	1.81	(1.24-2.64)	
Naldi *et al.*[273]	526 cases 512 controls	**Lifetime recreational sun exposure**				
		Intermediate vs. low	AG	1.8	(1.3-2.6)	<0.001
		High vs. low	AG	1.6	(1.1-2.3)	
Suárez-Varela *et al.*[275]	Men (number unclear)	**Sun exposure during outdoor activities (hours/day)**				
		0.005-0.12 vs. < 0.005	A	0.7[a]	(0.5-1.3)	
		> 0.12 vs. < 0.005	A	1.7[a]	(1.1-2.8)	
		Sun exposure during holidays (hours/day)				
		0.16-0.26 vs. < 0.16	A	0.8[a]	(0.5-1.16)	
		0.26-0.40 vs. < 0.16	A	1.3[a]	(0.7-2.6)	
		> 0.40 vs. < 0.16	A	2.1[a]	(1.2-3.9)	
	Women (number unclear)	**Sun exposure during outdoor activities (hours/day)**				
		0.005-0.12 vs. < 0.005	A	1.4[a]	(0.5-3.8)	
		> 0.12 vs. < 0.005	A	2.1[a]	(0.9-4.7)	
		Sun exposure during holidays (hours/day)				
		0.16-0.26 vs. < 0.16	A	1.7[a]	(0.7-4.3)	
		0.26-0.40 vs. < 0.16	A	1.2[a]	(0.6-2.5)	
		> 0.40 vs. < 0.16	A	1.4[a]	(0.6-3.2)	

Table 9.6 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Gallagher *et al.*[271]	186 cases 326 controls	**Lifetime mean recreational sun exposure (hours/week in summer)**				
		2.8-5.5 vs. < 2.8	AGI	0.9	(0.5-1.7)	0.03
		5.6-8.4 vs. < 2.8	AGI	0.6	(0.3-1.3)	
		≥ 8.5 vs. < 2.8	AGI	0.4	(0.1-1.0)	

A: Adjusted for general factors, such as age and sex
B: Adjusted for region of residence
C: Adjusted for skin sensitivity to sunlight
D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer
F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour
H: Adjusted for other environmental risk factors
I: Adjusted for ethnicity
* *Includes all non-melanoma skin cancers*

9.2.4.4 Total sun exposure

In some studies no distinction has been made between occupational and recreational sun exposure, and the effect of total sun exposure on BCC has been investigated. In a cohort study in the United States,[264] there was evidence that the risk of BCC was increased when comparing women who spent more than eight hours per week outside in summer and used sunscreen with those who did not spend more than eight hours per week outside, with a hazard ratio of 1.37 (1.11-1.69).

There was little evidence of an association between cumulative exposure to UVB radiation and BCC in a cross-sectional study in the United States.[269] Comparing men with above median exposure with those with below median exposure, the OR for BCC was 0.69 (0.31-1.53). Comparing men in the upper quartile of UVB radiation exposure with men in the lower three quartiles resulted in an OR for BCC of 1.11 (0.50-2.44).

Little evidence of an association between lifetime total sun exposure and BCC was found in a case-control study in Canada.[271] When compared with a mean value of less than 11.5 hours per week of sun exposure in summer, between 11.5 and 18.9 hours per week of summer exposure had an OR for BCC of 1.3 (0.8-2.2), between 19 and 27.9 hours per week of summer exposure had an OR of 1.2 (0.7-2.2), and 28 hours per week or more of summer exposure had an OR of 1.3 (0.7-2.4). The p-value for trend was greater than 0.05.

Data from a case-control study in Western Australia was used to investigate the association between total sun exposure and BCC.[266] There was some evidence that the risk of BCC on the head and neck decreased with increasing sun exposure to the site. When compared with 28,600 hours or less of sun exposure to the head and neck, between 28,600 hours and 44,500 hours of exposure gave an OR for BCC of 0.95 (0.45-1.97), between 44,500 hours and 65,000 hours of exposure gave an OR for BCC of 0.90 (0.42-1.89), and 65,000 hours of exposure or more gave an OR for BCC of 0.38 (0.15-0.97). The p-value for trend was 0.07. In contrast, there was evidence that the risk of BCC on the trunk increased with increasing sun exposure to the site. When compared with no sun exposure to the trunk, between 100 and 5,200 hours of exposure gave an OR for BCC of 0.62 (0.24-1.59), between 5,200 hours and 13,000 hours of exposure gave an OR for BCC of 1.03 (0.51-2.09), and 13,000 hours or more of exposure gave an OR for BCC of 2.39 (1.18-4.83). The p-value for trend was 0.01. There was weak evidence that the risk of BCC on the limbs was decreased by increasing sun exposure to the site. When compared with 19,200 hours of exposure or less, between 19,200 hours and 28,600 hours of exposure had an OR for BCC of 1.21 (0.41-3.62), between 28,600 hours and 43,700 hours of exposure had an OR for BCC of 0.78 (0.25-2.37), and 43,700 hours or more of exposure had an OR for BCC of 0.43 (0.13-1.44). The p-value for trend was 0.15. Using sun exposure to each site as a continuous exposure

Table 9.7: Summary of published associations between total sun exposure and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Hunter *et al.*[264]	73,366 women	**Regularly spending at least 8 hours per week outdoors in summer**				
		Yes with sunscreen vs. no	ABCDFG	1.37[a]	(1.11-1.69)	
Vitasa *et al.*[269]	808 men	**Cumulative exposure to ultraviolet-B**				
		Above median vs. below median	ACEG	0.69	(0.31-1.53)	
		Top quartile vs. lower three quartiles	ACEG	1.11	(0.50-2.44)	
Gallagher *et al.*[271]	186 cases 326 controls	**Lifetime mean sun exposure (hours/week in summer)**				
		11.5-18.9 vs. < 11.5	AGI	1.3	(0.8-2.2)	> 0.05
		19.0-27.9 vs. < 11.5	AGI	1.2	(0.7-2.2)	
		≥ 28 vs. < 11.5	AGI	1.3	(0.7-2.4)	
Kricker *et al.*[266]	69 cases 307 controls	**Sun exposure on head and neck (hours)**				
		28,600-44,500 vs. 0-28,600	AC	0.95	(0.45-1.97)	0.07
		44,500-65,000 vs. 0-28,600	AC	0.90	(0.42-1.89)	
		> 65,000 vs. 0-28,600	AC	0.38	(0.15-0.97)	
	82 cases 263 controls	**Sun exposure on trunk (hours)**				
		100-5,200 vs. none	AC	0.62	(0.24-1.59)	0.01
		5,200-13,000 hours vs. none	AC	1.03	(0.51-2.09)	
		> 13,000 vs. none	AC	2.39	(1.18-4.83)	
	41 cases 130 controls	**Sun exposure on limbs (hours)**				
		19,200-28,600 vs. 0-19,200	AC	1.21	(0.41-3.62)	0.15
		28,600-43,700 vs. 0-19,200	AC	0.78	(0.25-2.37)	
		≥ 43,700 vs. 0-19,200	AC	0.43	(0.13-1.44)	

A: Adjusted for general factors, such as age and sex, B: Adjusted for region of residence, C: Adjusted for skin sensitivity to sunlight, D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer, F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour, H: Adjusted for other environmental risk factors, I: Adjusted for ethnicity
[a] Hazard ratio

measure, quadratic models showed that risk of BCC increased up to a total exposure of approximately 35,000 hours, and then fell.

Table 9.7 summarises the estimated associations between total sun exposure and BCC described above.

9.2.4.5 Sun exposure before the age of 20

Evidence of an association between intermittency of sun exposure between the ages of 15 and 19 and BCC has been found in a case-control study in Western Australia.[279] Intermittency was measured as the proportion of days exposed to sunlight on non-working days compared with the total number of days exposed. Between the ages of 15 and 19, comparing 100% intermittency with less than 40% gave an OR for BCC of 3.86 (1.93-7.75). Comparing between 59% and 99% intermittency with less than 40% gave an OR for BCC of 1.82 (1.01-3.28), while comparing between 41% and 58% intermittency with less than 40% gave an OR for BCC of 1.49 (0.88-2.52). The p-value for trend was less than 0.001, providing strong evidence that the risk of BCC increases with increasing intermittency of sun exposure between the ages of 15 and 19.

In a case-control study of Canadian men,[271] some evidence of a harmful association between sunburns between the ages of 5 and 15 and BCC was found. When comparing frequent or severe burns with no burns, the OR for BCC was 1.6 (1.0-2.7). There was little evidence of an association with less severe burning and BCC, however. When compared with no burns between the ages of 5 and 15, moderate burns had an OR for BCC of 1.3 (0.8-2.1) and rare or mild burns had an OR for BCC of 0.8 (0.5-1.4). Having sunburn pain for two days or more, twice a year or more between the ages of 5 and 15 when compared with no pain gave an OR for BCC of 4.5 (1.7-12.3), while having sunburn pain for two days or more once a year had an OR for BCC of 1.7 (0.9-3.4) when compared with no pain. Mean recreational sun exposure of more than 12.5 hours per week in summer before the age of 19 years showed evidence of a harmful association with BCC when compared with less than 3.8 hours per week, with an OR for BCC of 2.6 (1.1-6.5). There was little evidence of an association between lower categories of sun exposure at under 19 years of age and BCC. When compared with less than 3.8 hours per week in summer, between 3.8 and 7.4 hours per week had an OR for BCC of 1.1 (0.6-2.0), and between 7.5 and 12.4 hours per week had an OR for BCC of 1.4 (0.7-3.0). There was, however, evidence that increasing hours of sun exposure per week in summer before the age of 19 increased the risk of BCC, as the p-value for trend was 0.03.

A case-control study in Southern Europe[272] found evidence of an increased risk of BCC among people who experienced a first sunburn before the age of 15. Compared with people whose first sunburn was after the age of 15, the OR for BCC was 1.65 (1.16-2.36). In a separate publication,[281] there was evidence of a harmful association between sun exposure at the beach

during childhood holidays and BCC. When compared with no sun exposure at the beach in childhood, less than 197 hours of exposure gave an OR for BCC of 1.05 (0.78-1.41), between 197 and 714 hours of exposure gave an OR for BCC of 1.23 (0.93-1.64), between 715 and 2,079 hours of exposure gave an OR for BCC of 1.10 (0.82-1.48), and more than 2,079 hours of exposure gave an OR for BCC of 1.43 (1.09-1.89). The p-value for trend was 0.005, providing strong evidence that the risk of BCC increases as the amount of sun exposure at the beach in childhood increases.

Analyses from the Health Professionals Follow-Up Study found evidence of an association between time spent outdoors in a swimming costume as a teenager in summer and BCC.[267] Compared with people who were outdoors less than once a week in summer in a swimming costume, people who were outdoors once a week had a hazard ratio for BCC of 1.30 (1.14-1.47), people who were outdoors twice a week had a hazard ratio for BCC of 1.34 (1.19-1.52), people who were outdoors several times a week had a hazard ratio for BCC of 1.36 (1.22-1.52), and people who were outdoors in a swimming costume every day had a hazard ratio for BCC of 1.42 (1.24-1.63). The p-value for trend was less than 0.0001, providing strong evidence that the risk of BCC increases as the amount of time spent outdoors in a swimming costume in summer as a teenager increases.

An Italian case-control study[273] found evidence of an association between intermediate amounts of recreational sun exposure before the age of 20 and BCC, with an OR for BCC of 1.9 (1.2-3.1) when compared with low amounts of exposure. There was little evidence of an effect of high levels of sun exposure as the OR for BCC was 1.0 (0.6-1.6) when compared with low levels of sun exposure. Evidence was found of a harmful association between severe sunburns before the age of 15 and BCC. When compared with no severe sunburns, one severe sunburn gave an OR of 2.8 (1.9-4.1), while two or more gave an OR of 3.9 (1.6-9.2). The p-value for trend was less than 0.001, which provides strong evidence of an increasing risk of BCC with increasing numbers of sunburns before the age of 15.

The estimated associations between sun exposure before age 20 and BCC described above are summarised in Table 9.8.

9.2.4.6 Summary

The evidence for an association between sun exposure and BCC is mixed. There is some evidence that sunburn is related to BCC, and that the risk of BCC increases as the frequency of sunburns increases. The evidence for an effect of occupational or total sun exposure is less convincing. There is little evidence that increasing occupational exposure increases the risk of BCC, and in fact increasing total sun exposure to the head or limbs may protect against BCC at these sites. The evidence for recreational exposure points to an increased risk of BCC with

Table 9.8: Summary of published associations between sun exposure before the age of twenty and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Kricker *et al.*[279]	201 cases 700 controls	**Intermittency of sun exposure (ages 15-19)**				
		41-58% vs. 0-40%	ACD	1.49	(0.88-2.52)	<0.001
		59-99% vs. 0-40%	ACD	1.82	(1.01-3.28)	
		100% vs. 0-40%	ACD	3.86	(1.93-7.75)	
Gallagher *et al.*[271]	199 cases 406 controls	**Sunburn (ages 5-15)**				
		Rare or mild burns vs. never burned	AGI	0.8	(0.5-1.4)	
		Moderate burns vs. never burned	AGI	1.3	(0.8-2.1)	
		Frequent or severe burns vs. never burned	AGI	1.6	(1.0-2.7)	
	226 cases 406 controls	**≥ 2 days of sunburn pain (ages 5-15)**				
		Once per year vs. never	AGI	1.7	(0.9-3.4)	
		2+ per year vs. never	AGI	4.5	(1.7-12.3)	
	216 cases 387 controls	**Mean recreational sun exposure per year before age 20 (hours/week in summer)**				
		3.8-7.4 vs. < 3.8	AGI	1.1	(0.6-2.0)	0.03
		7.5-12.4 vs. < 3.8	AGI	1.4	(0.7-3.0)	
		≥ 12.5 vs. < 3.8	AGI	2.6	(1.1-6.5)	
Zanetti *et al.*[272]	1,549 cases 1,795 controls	**Age at first sunburn**				
		≤ 15 years vs. > 15 years or never	ABCG	1.65	(1.16-2.36)	
Rosso *et al.*[281]	1,549 cases 1,795 controls	**Childhood sun exposure at the beach on holidays (hours)**				
		< 197 vs. none	ABCG	1.05	(0.78-1.41)	0.005
		197-714 vs. none	ABCG	1.23	(0.93-1.64)	
		715-2,079 vs. none	ABCG	1.10	(0.82-1.48)	
		> 2,079 vs. none	ABCG	1.43	(1.09-1.89)	

Table 9.8 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
van Dam *et al.*[267]	248,813 person-years 2,705 cases	**Frequency outdoors in a swimming costume as a teenager in summer**				
		Once a week vs. less than once a week	ABCGI	1.30	(1.14-1.47)	<0.0001
		Twice a week vs. less than once a week	ABCGI	1.34	(1.19-1.52)	
		Several times a week vs. less than once a week	ABCGI	1.36	(1.22-1.52)	
		Daily vs. less than once a week	ABCGI	1.42	(1.24-1.63)	
Naldi *et al.*[273]	528 cases 512 controls	**Severe sunburns before age 15**				
		One vs. none	AG	2.8	(1.9-4.1)	<0.001
		≥ 2 vs. none	AG	3.9	(1.6-9.2)	
		Recreational sun exposure before age 20				
		Intermediate vs. low	AG	1.9	(1.2-3.1)	0.001
		High vs. low	AG	1.0	(0.6-1.6)	

A: Adjusted for general factors, such as age and sex
B: Adjusted for region of residence
C: Adjusted for skin sensitivity to sunlight
D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer
F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour
H: Adjusted for other environmental risk factors
I: Adjusted for ethnicity
[a] *Rate ratio*
[b] *Hazard ratio*
[c] *Includes all non-melanoma skin cancers*

increasing recreational exposure, but this association has not been found consistently. Sunburn and recreational sun exposure before the age of 20 also appear to be risk factors for BCC. The evidence suggests that the risk of BCC increases with increasing frequency of sunburn before the age of 15 and increasing recreational sun exposure in childhood.

Several reasons have been proposed for this observed association between recreational sun exposure, sunburn and BCC, but a lack of association with occupational or total sun exposure. It has been suggested that a dose-response relationship exists between sun exposure and BCC, where at higher levels of exposure the risk of BCC levels out, and possibly even drops.[266, 269] Occupational sun exposure may therefore not show an association with BCC, as sun exposure is high. Kricker *et al.*[266] suggested two explanations for this dose-response relationship. Mutations due to exposure to UV radiation may be very common in sun exposed skin.[283, 284] As sun exposure increases, these mutations may lead to cells that would become BCCs to become non-viable. The second hypothesis proposed was that the mutations caused by sun exposure could increase the immunogenicity of the basal cells (the degree to which the cells induce an immune response), and therefore make them more susceptible to destruction by the immune system. On the other hand, this dose-response relationship may not truly exist and may be due to people who are aware of their high risk of developing BCCs (because of previous skin cancers or benign skin lesions) reducing their sun exposure. People with high levels of exposure would then be at lower risk of BCC.[266]

The pattern of sun exposure may also be important in determining the risk of BCC.[266] Recreational sun exposure and sunburn are both indicators of intermittent sun exposure, and this intermittency may be important, particularly in childhood.[279] Rosso *et al.*[281] suggested that outdoor work, which indicates more consistent sun exposure, allows the skin to develop sun protection by tanning and thickening of the external layers. This protection does not develop to the same extent with recreational exposure, and therefore BCC is caused when UV radiation reaches the basal cells as directly as possible, in the absence of any natural skin protection. This hypothesis also supports the dose-response curve with a levelling of risk at high levels of sun exposure.

Finally, it has been suggested that sun exposure in childhood is an important risk factor for BCC,[271, 279, 281] and that the skin in childhood is more sensitive to the effects of sun exposure that lead to BCCs developing in later life.[271] Armstrong *et al.*[280] suggested that the potential for skin cancer is determined by sun exposure in childhood. Adult sun exposure then determines how much of this potential is realised.

9.2.5. Previous benign sun-related skin conditions

Associations between BCC and benign sun-related skin conditions, such as actinic keratoses,

actinic cheilitis, solar lentigines, melanocytic naevi and actinic elastosis have been found. Actinic keratoses are skin lesions that begin as flat scaly areas, but later develop a hard, wart-like surface.[285] If an actinic keratosis occurs on the lip, it is called actinic cheilitis.[286] Solar lentigines, also known as age spots or liver spots, are flat spots with increased pigmentation.[287] Melanocytic naevi are also known as moles and are clusters of pigmented skin cells.[288] Actinic elastosis is premature aging of the skin due to sun exposure. Some of the characteristics of actinic elastosis are inelasticity and wrinkling of the skin and skin dryness with fine scaling.[289] Telangiectasias are abnormally dilated blood vessels. Facial telangiectasia is associated with age, sun exposure and alcohol consumption.[290]

A cohort study in Queensland, Australia found evidence of a harmful association between actinic keratoses and BCC.[282] When compared with people with no actinic keratoses on the face, people with between one and five had an OR for BCC of 3.9 (1.9-8.0), people with between six and 20 had an OR for BCC of 5.6 (2.3-13.3), and people with more than 20 had an OR for BCC of 10.0 (3.5-28.2). There was also evidence that facial telangiectasia had a harmful association with BCC. Compared with people with no facial telangiectasia, people with mild telangiectasia had an OR for BCC of 2.3 (1.1-4.7), people with moderate telangiectasia had an OR for BCC of 2.9 (1.2-7.1), and people with severe telangiectasia had an OR for BCC of 7.3 (2.1-26.0). There was evidence of a harmful association between actinic elastosis of the neck and BCC. Compared with people with no elastosis of the neck, people with mild to moderate actinic elastosis had and OR for BCC of 3.7 (1.6-8.3) and people with severe actinic elastosis of the neck had an OR for BCC of 3.6 (1.3-9.8). There was also evidence of a harmful association between the number of solar lentigines on the hands and BCC. Compared with people with no solar lentigines on the hands, people with between one and 10 had an OR for BCC of 1.5 (0.8-2.5), people with between 11 and 20 had an OR for BCC of 2.9 (1.2-7.0), and people with more than 20 had an OR for BCC of 3.7 (1.2-11.7).

Evidence of a harmful association between childhood freckling and BCC was found in a cross-sectional study in the United States.[269] Comparing men who experienced freckling during childhood with men who did not, the OR for BCC was 3.66 (1.51-8.84).

Some evidence of associations between benign sun related skin conditions and BCC was found in a case-control study in Western Australia.[270] The majority of analyses adjusted for age, sex, age of arrival in Australia and Southern European ethnicity. Evidence of an increased risk of BCC with increased density of freckling in childhood on the arms was found. Comparing people who had scattered freckling on the arms during childhood with people who had no freckling on the arms, the OR for BCC was 1.37 (0.93-2.02). Comparing people with moderate to heavy freckling on the arms during childhood with people with none gave an OR for BCC of 1.63 (1.06-2.51). The p-value for trend was 0.02. There was also evidence of an increased risk of

BCC for people with four or more melanocytic naevi on the back. Comparing people with four or more melanocytic naevi on the back with people with three or less, the OR for BCC was 1.80 (1.22-2.65). Strong evidence was found for an increasing risk of BCC with increasing severity of actinic elastosis on the neck. Comparing people with mild solar elastosis on the neck with people with none, the OR for BCC was 1.85 (0.80-4.26). When compared with people with no actinic elastosis on the neck, people with moderate actinic elastosis had an OR for BCC of 2.75 (1.16-6.50), and people with severe actinic elastosis of the neck had an OR for BCC of 3.96 (1.58-9.93). The p-value for trend was less than 0.001.

A Canadian case-control study found evidence of a harmful association between freckling between the ages of five and 15 and BCC.[271] When compared with people who did not freckle between those ages, people who did had an OR for BCC of 1.8 (1.2-2.5).

Analyses from the Health Professionals Follow-Up Study found evidence of an association between the number of melanocytic naevi on the forearms and risk of BCC.[267] Compared with people with no melanocytic naevi on the forearms, people with one or two had a hazard ratio for BCC of 1.27 (1.14-1.41), people with between three and five had a hazard ratio for BCC of 1.29 (1.11-1.50), and people with six or more had a hazard ratio for BCC of 1.20 (1.00-1.45). The p-value for trend was less than 0.0001, providing strong evidence that the risk of BCC increases with increasing numbers of melanocytic naevi on the forearms.

In an Italian case-control study,[273] an OR for BCC from many solar lentigines compared with none of 1.6 (1.1-2.1) was found, although there was little evidence of an association when comparing few solar lentigines with none (OR=1.1 (0.8-1.4)). The p-value for trend was 0.002, providing strong evidence of an increasing risk of BCC with increasing numbers of solar lentigines. There was evidence of a harmful association between actinic keratoses and BCC. Comparing people with actinic keratoses with those with none, the OR for BCC was 2.8 (2.0-4.0). There was little evidence of an association between freckles and BCC, with an OR for BCC of 0.8 (0.6-1.1) when comparing people with freckles with people without freckles. There was evidence of a harmful association between melanocytic naevi on the upper limbs and BCC. Comparing people with between one and five melanocytic naevi on the upper limbs with those with none, the OR for BCC was 1.7 (1.2-2.3), while comparing those with six or more with those with none gave an OR of 1.6 (1.1-2.4). The p-value for trend was 0.002, providing strong evidence for an increase in risk for BCC with increasing numbers of melanocytic naevi on the upper limbs.

In a Yugoslavian case-control study, the OR for BCC for people who freckled before the age of 15 was 2.65 (1.29-5.46),[274] which provides evidence for a harmful effect. The reference category for this analysis was unclear.

Table 9.9: Summary of published associations between benign sun-related skin conditions and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
Green *et al.*[282]	1,770 subjects	**Number of actinic keratoses on the face**				
		1-5 vs. none	A	3.9	(1.9-8.0)	
		6-20 vs. none	A	5.6	(2.3-13.3)	
		> 20 vs. none	A	10.0	(3.5-28.2)	
		Telangiectasia of the face				
		Mild vs. none	A	2.3	(1.1-4.7)	
		Moderate vs. none	A	2.9	(1.2-7.1)	
		Severe vs. none	A	7.3	(2.1-26.0)	
		Actinic elastosis of the neck				
		Mild/moderate vs. none	A	3.7	(1.6-8.3)	
		Severe vs. none	A	3.6	(1.3-9.8)	
		Number of solar lentigines on the hands				
		1-10 vs. none	A	1.5	(0.8-2.8)	
		11-20 vs. none	A	2.9	(1.2-7.0)	
		> 20 vs. none	A	3.7	(1.2-11.7)	
Vitasa *et al.*[269]	808 men	**Freckling in childhood**				
		Yes vs. no	ACDG	3.66	(1.51-8.84)	
Kricker *et al.*[270]	226 cases 1,021 controls	**Freckling on arms in childhood**				
		Scattered vs. none	ACEI	1.37	(0.93-2.02)	0.02
		Moderate/heavy vs. none	ACEI	1.63	(1.06-2.51)	
		Number of melanocytic naevi on the back				
		≥ 4 vs. 0-3	ACEI	1.80	(1.22-2.65)	
		Actinic elastosis of the neck				
		Mild vs. none	ACEI	1.85	(0.80-4.26)	<0.001
		Moderate vs. none	ACEI	2.75	(1.16-6.50)	
		Severe vs. none	ACEI	3.96	(1.58-9.93)	
Gallagher *et al.*[271]	226 cases 406 controls	**Freckling (age 5-15)**				
		Yes vs. no	AGI	1.8	(1.2-2.5)	

Table 9.9 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI	P-value for trend
van Dam *et al.*[267]	204,751 person-years	**Number of melanocytic naevi on the forearms**				
	2,203 cases	1-2 vs. none	ACGI	1.27[a]	(1.14-1.41)	<0.0001
		3-5 vs. none	ACGI	1.29[a]	(1.11-1.50)	
		≥ 6 vs. none	ACGI	1.20[a]	(1.00-1.45)	
Naldi *et al.*[273]	528 cases	**Solar lentigines**				
	512 controls	Few vs. none	AEG	1.1	(0.8-1.4)	0.002
		Many vs. none	AEG	1.6	(1.1-2.1)	
		Actinic keratosis				
		Yes vs. no	AEG	2.8	(2.0-4.0)	
		Freckles				
		Few/many vs. none	AEG	0.8	(0.6-1.1)	
		Melanocytic naevi on the upper limbs				
		1-5 vs. none	AEG	1.7	(1.2-2.3)	0.002
		≥6 vs. none	AEG	1.6	(1.1-2.4)	
Vlajinac *et al.*[274]	200 cases	**Freckles**				
	399 controls	Freckling before the age of 15	CDEGH	2.65	(1.29-5.46)	
Walther *et al.*[277]	146 cases	**Actinic cheilitis**				
	264 controls	Yes vs. no	ABDEG	7.1	(2.7-18.4)	
		Actinic keratosis				
		Any vs. none	ABDEG	2.7	(1.3-5.9)	
		Solar lentigines				
		Any vs. none	ABDEG	2.5	(1.2-5.3)	
		Actinic elastosis				
		Any vs. none	ABDEG	0.1	(0.05-0.42)	

A: Adjusted for general factors, such as age and sex, B: Adjusted for region of residence, C: Adjusted for skin sensitivity to sunlight, D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer, F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour, H: Adjusted for other environmental risk factors, I: Adjusted for ethnicity
[a] Hazard ratio

A German case-control study[277] comparing people with actinic cheilitis with people without gave an OR for BCC 7.1 (2.7-18.4). Comparing people with actinic keratosis to those without gave an OR for BCC of 2.7 (1.3-5.9), and comparing people with solar lentigo to people without gave an OR for BCC of 2.5 (1.2-5.3). The presence of actinic elastosis appeared to be protective against BCC, with an OR for BCC of 0.1 (0.05-0.42) when comparing people with actinic elastosis to people without.

The results described above are summarised in Table 9.9. Actinic keratoses appear to be associated with BCC, and may suggest that BCC is related to total sun exposure.[280] Telangiectasia also appears to have an association with BCC, although it was only considered in two studies. There is some evidence that solar lentigines are associated with BCC, although one study found little evidence of an association.[279] The evidence for an association between actinic elastosis and BCC is contradictory, with two studies reporting harmful effects[270, 282] and one study reporting a protective effect.[277] Freckling and the presence of melanocytic naevi appear to be associated with BCC, although the observed associations for melanocytic naevi may be due to misclassification of solar lentigines.[267, 270] It has been suggested that BCC and malignant melanoma are aetiologically similar, as freckles and melanocytic naevi are well known risk factors for malignant melanoma.[270, 271, 273, 274, 279]

9.2.6. Sun protection

9.2.6.1 Hat use

Use of a hat to protect against skin cancer was investigated by in a case-control study in Western Australia.[279] Evidence of an increased risk of BCC was found for people using hats more than half the time for between 10 and 19 years when compared with people who reported using hats less than half the time, with an OR for BCC of 2.90 (1.12-7.50). For use more than half the time for between one and nine years, and more than 20 years, the ORs for BCC were 2.46 (0.96-6.26) and 1.55 (0.69-3.47) respectively when compared with people who use a hat for less than half the time. Little evidence of an association between wearing a broad brimmed hat and BCC was found. When compared with people using a broad brimmed hat less than half the time, those using a broad brimmed hat more than half the time for between one and nine years had an OR for BCC of 1.15 (0.51-2.59), for between 10 and 19 years had an OR for BCC of 1.30 (0.50-3.41), and for 20 years or more had an OR for BCC of 0.77 (0.32-1.84). Considering hat use in the 10 years prior to diagnosis, using a hat for between one and nine years when compared with never using a hat gave an OR for BCC of 2.45 (1.10-5.45). Using a hat for the full 10 years showed little evidence of an association with BCC, with an OR for BCC of 1.43 (0.74-2.79) when compared with never wearing a hat. Little evidence of an association between hat use between 11 and 30 years prior to diagnosis and BCC was found. Compared with people who never wore a hat, those who wore a hat for between one and nine years had an OR for BCC of 1.39 (0.61-

3.16), for between 10 and 19 years had an OR for BCC of 0.82 (0.37-1.83), and for 20 or more years had an OR for BCC of 1.07 (0.52-2.22). There was similarly little evidence of an association between broad brimmed hat use and BCC. Comparing people who ever used a broad brimmed hat in the 10 years before diagnosis with those who never used a broad brimmed hat gave an OR for BCC of 1.23 (0.63-2.38). Comparing people who ever used a broad brimmed hat between 11 and 30 years prior to diagnosis with people who never wore a broad brimmed hat in this time period gave an OR for BCC of 1.05 (0.53-2.10).

Some evidence of a protective effect of using a hat during occupational sun exposure on NMSC was found among men in a Spanish case-control study.[275] Using a hat between May and September for occupational exposure gave an OR for NMSC of 0.5 (0.3-0.7), while between October and April the OR for NMSC was 0.5 (0.3-0.8), both when compared with not using a hat. There was, however, evidence of a harmful effect of hat use during holidays on NMSC among men. When compared with men who did not wear a hat, men who wore a hat during holidays between October and April had an OR for NMSC of 2.0 (1.3-3.3). Little evidence of an effect of hat use during outdoor activities on NMSC was found. Comparing men who used a hat during outdoor activities between May and September with men who did not gave an OR for NMSC of 1.4 (0.8-2.5), and comparing men who wore a hat during outdoor activities between October and April with those who did not gave an OR for NMSC of 1.3 (0.7-2.0). There was weak evidence of an effect of hat use during holidays between May and September among men with an OR for NMSC of 0.7 (0.4-1.3).

There was little evidence of any effect of hat use on NMSC among women. The ORs for NMSC were 0.8 (0.4-1.7) for hat use during occupational exposure between May and September, and 0.91 (0.3-2.5) for hat use during occupational exposure between October and April, both when compared with not wearing a hat. The ORs for NMSC were 0.7 (0.2-3.3) for hat use during outdoor activities between May and September, and 0.7 (0.2-2.5) for hat use during outdoor activities between October and April, both when compared with not wearing a hat. The ORs for NMSC were 0.8 (0.3-1.7) for hat use during holidays between May and September, and 1.1 (0.6-2.0) for hat use during holidays between October and April, both when compared with not using a hat. The effect of hat use on BCC alone cannot be established from these analyses, as they include both SCC and BCC.

The estimated associations between hat use and BCC described above are summarised in Table 9.10.

Table 9.10: Summary of published associations between hat use and basal cell carcinoma.

Reference	Number	Exposure	Control for confounding	OR	95% CI
Kricker *et al.*[279]	71 cases 291 controls	**Use of any hat**			
		More than half the time for 1-9 years vs. less than half the time	AC	2.46	(0.96-6.26)
		More than half the time for 10-19 years vs. less than half the time	AC	2.90	(1.12-7.50)
		More than half the time for ≥ 20 years vs. less than half the time	AC	1.55	(0.69-3.47)
		Use of a broad brimmed hat			
		More than half the time for 1-9 years vs. less than half the time	AC	1.15	(0.51-2.59)
		More than half the time for 10-19 years vs. less than half the time	AC	1.30	(0.50-3.41)
		More than half the time for ≥ 20 years vs. less than half the time	AC	0.77	(0.32-1.84)
		Use of any hat in the 10 years prior to diagnosis			
		1-9 years vs. never	AC	2.45	(1.1-5.45)
		10 years vs. never	AC	1.43	(0.74-2.79)
		Use of any hat 11-30 years prior to diagnosis			
		1-9 years vs. never	AC	1.39	(0.61-3.16)
		10-19 years vs. never	AC	0.82	(0.37-1.83)
		≥ 20 years vs. never	AC	1.07	(0.52-2.22)
		Use of a broad brimmed hat in the 10 years prior to diagnosis			
		Ever vs. never	AC	1.23	(0.63-2.38)
		Use of a broad brimmed hat 11-30 years prior to diagnosis			
		Ever vs. never	AC	1.05	(0.53-2.10)
Suárez-Varela *et al.*[275]	Men (number unclear)	**Hat use during occupational exposure (May-September)**			
		Yes vs. no	A	0.5[a]	(0.3-0.7)
		Hat use during occupational exposure (October-April)			
		Yes vs. no	A	0.5[a]	(0.3-0.8)
		Hat use during outdoor activities (May-September)			
		Yes vs. no	A	1.4[a]	(0.8-2.5)

Table 9.10 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI
Suárez-Varela *et al.*[275]	Men (number unclear)	**Hat use during outdoor activities (October-April)**			
		Yes vs. no	A	1.3[a]	(0.7-2.0)
		Hat use during holidays (May-September)			
		Yes vs. no	A	0.7[a]	(0.4-1.3)
		Hat use during holidays (October-April)			
		Yes vs. no	A	2.0[a]	(1.3-3.3)
	Women (number unclear)	**Hat use during occupational exposure (May-September)**			
		Yes vs. no	A	0.8	(0.4-1.7)
		Hat use during occupational exposure (October-April)			
		Yes vs. no	A	0.91	(0.3-2.5)
		Hat use during outdoor activities (May-September)			
		Yes vs. no	A	0.7	(0.2-3.3)
		Hat use during outdoor activities (October-April)			
		Yes vs. no	A	0.7	(0.2-2.5)
		Hat use during holidays (May-September)			
		Yes vs. no	A	0.8	(0.3-1.7)
		Hat use during holidays (October-April)			
		Yes vs. no	A	1.1	(0.6-2.0)

A: Adjusted for general factors, such as age and sex, B: Adjusted for region of residence, C: Adjusted for skin sensitivity to sunlight, D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer, F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour, H: Adjusted for other environmental risk factors, I: Adjusted for ethnicity
[a] Includes both SCCs and BCCs

9.2.6.2 Sunscreen use

There is little evidence that sunscreen protects against skin cancer. A review of the evidence in 2000[291] identified four studies which investigated the relationship between BCC and sunscreen use.[264, 275, 278, 279]

A cohort study among women in the United States[264] found evidence that, among women who spent more than eight hours outdoors in summer, the risk of BCC was increased for women who used sunscreen than for those that did not use it. The hazard ratio for BCC in this situation was 1.43 (1.22-1.67).

A case-control study in Western Australia[279] found evidence that using sunscreen half the time or more for between one and nine years when compared with using sunscreen less than half the time increased the risk of a BCC, with an OR of 1.92 (1.17-3.13). When using sunscreen for half the time or more for ten years or more was compared with less than half the time there was little evidence of an association as the OR for BCC was 1.25 (0.82-1.90). There was little evidence of an effect of sunscreen use on BCC when used between 11 and 30 years prior to diagnosis. When comparing use for half the time or more for between one and nine years with use less than half the time, the OR for BCC was 1.20 (0.69-2.08). Using sunscreen for more than half the time for ten years or more gave an OR for BCC of 0.72 (0.40-1.28) when compared with sunscreen use less than half the time. Considering sunscreen use in the ten years prior to diagnosis, using sunscreen more than half the time for less than 10 years produced an OR for BCC of 1.77 (1.09-2.87) when compared with no sunscreen use in the 10 years prior to diagnosis. Using sunscreen half the time or more for the full 10 years gave an OR for BCC of 1.07 (0.69-1.67).

There was some evidence of an harmful association between sunscreen use and NMSC among men in a case-control study in Spain.[275] Comparing men who used sunscreen with those who did not, the OR for NMSC was 1.7 (0.91-3.3). Little evidence of an association between NMSC and sunscreen use was found in women. Comparing women who used sunscreen with those who did not, the OR for NMSC was 1.4 (0.7-2.5). The analyses included both SCC and BCC, and therefore the effect of sunscreen on BCC alone cannot be determined.

The Nambour Skin Cancer and Actinic Eye Disease Trial was an RCT investigating the effect of sunscreen use and beta-carotene supplementation on NMSCs in a population from Queensland, Australia.[278] In this RCT, subjects were allocated to daily sunscreen use, or discretionary sunscreen use. Little evidence of effect of sunscreen use on BCC was found. The rate ratio for the number of people with new BCCs was 1.03 (0.73-1.46) when comparing the sunscreen intervention group with the control group, while the rate ratio for the total number of new

BCCs was 1.05 (0.82-1.34) when comparing the sunscreen intervention group with the control group.

Data from the Nambour Trial was also used to investigate the effect of sunscreen use on repeated occurrence of BCC.[292] Using a Cox proportional hazards model for a survival analysis of the time to first BCC, the hazard ratio for BCC when comparing the intervention group with the control group was 1.03 (0.77-1.38). Three different models were used for the multiple failure time survival analysis; the Andersen-Gill model,[293] the Wei-Lin-Weissfeld model,[251] and the Prentice-Williams-Peterson model.[294] When all failures were analysed simultaneously, the Andersen-Gill model gave a hazard ratio for BCC of 0.90 (0.66-1.23), the Wei-Lin-Weissfeld model gave a hazard ratio for BCC of 0.89 (0.65-1.24), and the Prentice-Williams-Peterson model gave a hazard ratio for BCC of 0.91 (0.72-1.15), all when comparing the intervention group with the control group. When the analysis was stratified by event episodes, the estimated hazard ratio for the second occurrence of BCC from the Wei-Lin-Weissfeld model was 0.70 (0.43-1.16) when comparing the intervention arm with the control arm, and from the Prentice-Williams-Peterson model the estimated hazard ratio for the second occurrence of BCC was 0.71 (0.43-1.17) when comparing the intervention arm with the control arm. For the third occurrence of BCC, the estimated hazard ratio from the Wei-Lin-Weissfeld model was 0.59 (0.27-1.28) when comparing the intervention arm with the control arm, and the estimated hazard ratio from the Prentice-Williams-Peterson model was 0.67 (0.31-1.44) when comparing the intervention arm with the control arm..

The estimated associations between sunscreen use and BCC described above are summarised in Table 9.11.

Some evidence for a protective effect of sunscreen use comes from studies and trials for the prevention of actinic keratoses. In a trial in Australia, 588 subjects with between one and thirty actinic keratoses at baseline were randomised to daily sunscreen, or daily use of a cream without the sun protective active agents.[295] The trial found evidence that the risk of new actinic keratoses in the sunscreen group was reduced, compared with the control group, with a risk ratio for actinic keratosis of 0.62 (0.54-0.71).

Evidence of a protective effect of sunscreen use on actinic keratoses was found in an RCT conducted in the United States.[296] In total, 53 individuals joined the trial, with 26 randomised to the control group and 27 randomised to the sunscreen group. The participants were followed up for two years. Over that time, the rate of actinic keratosis formation per year in the sunscreen group was 51% (p=0.023) of the rate per year in the control group.

Table 9.11: Summary of published associations between sunscreen use and basal cell carcinoma

Reference	Number	Exposure	Control for confounding	OR	95% CI
Hunter *et al.*[264]	73,366 women	**Sunscreen use when regularly outside for > 8 hours/week in summer**			
		Yes vs. no	ABCDFG	1.43[a]	(1.22-1.67)
Kricker *et al.*[279]	192 cases	**Use of SPF 10+ sunscreen**			
	700 controls	More than half the time for 1-9 years vs. less than half the time	AC	1.92	(1.17-3.13)
		More then half the time for ≥ 10 years vs. less than half the time	AC	1.25	(0.82-1.90)
		Use of SPF 10+ sunscreen 11-30 years before diagnosis			
		More than half the time for 1-9 years vs. less than half the time	AC	1.20	(0.69-2.08)
		More than half the time for ≥ 10 years vs. less than half the time	AC	0.72	(0.40-1.28)
		Use of SPF 10+ sunscreen in the 10 years prior to diagnosis			
		More than half the time for 1-9 years vs. less than half the time	AC	1.77	(1.09-2.87)
		More than half the time for 10 years vs. less than half the time	AC	1.07	(0.69-1.67)
Suárez-Varela *et al.*[275]	Men	**Use of sunscreen**			
	(number unclear)	Yes vs. no	A	1.7[b]	(0.9-3.3)
	Women	**Use of sunscreen**			
	(number unclear)	Yes vs. no	A	1.4[b]	(0.7-2.5)
Green *et al.*[278]	1,621 subjects	**Number of participants with incident BCCs**			
		Intervention vs. control	-	1.03[c]	(0.73-1.46)
		Total number of incident BCCs			
		Intervention vs. control	-	1.05[c]	(0.82-1.34)
Pandeya *et al.*[292]	1,621 subjects	**Time to first BCC (Cox proportional hazards model)**			
		Intervention vs. control	-	1.03[a]	(0.77-1.38)
		Time to repeated BCCs (Andersen-Gill model)			
		Intervention vs. control	-	0.90[a]	(0.66-1.23)
		Time to repeated BCCs (Wei-Lin-Weissfeld model)			
		Intervention vs. control	-	0.89[a]	(0.65-1.24)
		Time to repeated BCCs (Prentice-Williams-Peterson model)			
		Intervention vs. control	-	0.91[a]	(0.72-1.15)
		Time to 2nd BCC (Wei-Lin-Weissfeld model)			
		Intervention vs. control	-	0.70[a]	(0.43-1.16)

Table 9.11 continued.

Reference	Number	Exposure	Control for confounding	OR	95% CI
Pandeya *et al.*[292]	1,621 subjects	**Time to 2nd BCC (Prentice-Williams-Peterson model)**			
		Intervention vs. control	-	0.71[a]	(0.43-1.17)
		Time to 3rd BCC (Wei-Lin-Weissfeld model)			
		Intervention vs. control	-	0.59[a]	(0.27-1.28)
		Time to 3rd BCC (Prentice-Williams-Peterson model)			
		Intervention vs. control	-	0.67[a]	(0.31-1.44)

A: Adjusted for general factors, such as age and sex
B: Adjusted for region of residence
C: Adjusted for skin sensitivity to sunlight
D: Adjusted for sun exposure
E: Adjusted for benign sun-related skin conditions and/or previous non-melanoma skin cancer
F: Adjusted for sun protection, such as use of sunscreen
G: Adjusted for phenotypic characteristics, such as eye colour and hair colour
H: Adjusted for other environmental risk factors
I: Adjusted for ethnicity
[a] *Hazard ratio*
[b] *Includes both SCCs and BCCs*
[c] *Rate ratio*

Data from the Nambour Trial were used to assess the effect of sunscreen use on actinic keratoses.[297] Although actinic keratoses increased over the period of the trial in both randomised groups, between 1992 and 1994 the increase in the number of actinic keratoses on the whole body in the sunscreen group was 76% (62%-94%) of the increase in actinic keratoses on the whole body in the control group. Similar results were found for actinic keratoses occurring on the sites of sunscreen application (head, neck, arms and back of hands), with the increase in actinic keratoses in the intervention group being 78% (64%-96%) of the increase in the control group. Little evidence of an effect of sunscreen use on the number of actinic keratoses was found between 1994 and 1996. The increase in the number of actinic keratoses on the whole body in the sunscreen group was 95% (75%-119%) of the increase in actinic keratoses on the whole body in the control group. The increase in actinic keratoses occurring on the sites of sunscreen application in the intervention group was 94% (75%-119%) of the increase in actinic keratoses on the sunscreen application sites in the control group. This lack of evidence of an association between 1994 and 1996 may be due to an increase in the amount of sunscreen used by subjects in the control arm.

9.2.6.3 Summary

There is very little evidence that sun protection behaviours such as wearing a hat or using sunscreen protects against BCC. For hat use, the only protective associations found were in a study which included both BCC and SCC as the outcome.[275] Squamous cell carcinoma may have a very different relationship with sun exposure than BCC, and this observed protective effect may be due to protection against SCC. In the study considering only BCC as the outcome, there was some evidence that hat use was harmful.[279] The effect of sunscreen on BCC was also unclear. In the observational studies, sunscreen appeared to have a harmful effect on BCC. Results from the Nambour trial showed little evidence of an effect of sunscreen on first BCC, but there may a protective effect of sunscreen on repeated occurrences of BCC. A review of the evidence for the effect of sunscreen on skin cancer conducted in 2000 *"concluded that there is inadequate evidence in humans for a cancer-preventive effect of topical use of sunscreen formulations against cutaneous malignant melanoma and basal-cell carcinoma of the skin"*,[291] and little evidence has been found to suggest that this conclusion should be changed in relation to BCC.

One possible explanation for the harmful associations seen in the observational studies is that people who are aware of a higher risk status for BCC (because of a previous skin cancer or benign skin lesion) start protecting themselves against sun exposure by using hats and applying sunscreen.[264, 279] This change of behaviour in high risk people would then lead to an apparent harmful effect of hat use or sunscreen on BCC. Any specific harmful effect of applying sunscreen seems unlikely because similar harmful effects are seen for hat use.[279]

The estimated associations reported in the observational studies are likely to be subject to

unmeasured confounding.[264, 280, 291] For example, people who use sunscreen may be more likely to have light hair and eye colour, or to have skin that is more sensitive to sun exposure. The use of sunscreen may lead to longer sun exposure, and to wearing less clothing for protection from sunlight.

9.3. Summary

The literature described above shows some evidence that UV radiation is a cause of BCC, although the relationship may not be straightforward. Some methodological issues should also be considered when evaluating the evidence.

For all observational studies described in this chapter, exposure measurement error is likely. It is very difficult to accurately measure a participant's sun exposure history, and many of the studies required participants to estimate their exposure many years in the past. In addition, recall bias in the case-control studies may have caused cases to report higher levels of sun exposure than controls. In this situation, associations between sun exposure and BCC may be overestimated.

A further possibility for all studies is outcome misclassification, especially in studies in which outcomes were classified according to subject self-report or clinical diagnosis, rather than histological diagnosis. Many of the studies, however, classified outcomes based on histological data to minimise this problem.

The lack of effect of sunscreens in protecting against BCC is an area that requires further research. The data of the Nambour Trial[278, 298] provide an opportunity for further analysis. In randomised controlled trials, while patients are allocated to receive treatment or not, they often do not comply with their assignation. The Nambour Trial Group started collecting data on sunscreen compliance in 1996, and these data are useful in estimating the efficacy of sunscreen on BCC. This analysis is presented in Chapter 10.

BLANK IN ORIGINAL

Chapter 10.

Estimating treatment efficacy in randomised controlled trials with departures from randomly allocated treatment: the effect of sunscreen use on time to first basal cell carcinoma

10.1. Introduction

In this chapter, the association between sunscreen use and time to first basal cell carcinoma (BCC) is investigated allowing for departures from the randomly assigned treatment regime. Data from the Nambour Skin Cancer and Actinic Eye Disease Trial (Nambour Trial) are used, which contain repeated measures of sunscreen use over almost six years. A rank preserving structural failure time model (RPSFTM) is used to relate the total time using sunscreen throughout the trial to the possibly counterfactual failure time that would have been observed had the subject used no sunscreen. The estimated acceleration parameter from the RPSFTM, which describes how fast lifetime is used up by using sunscreen, is converted into a hazard ratio, which is a more commonly used parameter in survival analyses. The analysis is carried out for two groups; the complete cases, which includes all subjects who did not drop out of the study before October 2002, and the complete cohort, which includes all subjects who entered the trial in 1992.

10.2. Methods

10.2.1. The Nambour Trial

The design of the Nambour Trial has been described in detail by Green, Battistutta, Hart *et al.*,[298] and so will be described more briefly here. It is a 2x2 factorial trial, aimed at assessing the effect of sunscreen use and beta-carotene supplementation on the incidence of squamous cell carcinoma (SCC) and BCC in a community in Queensland. The baseline survey for the trial was carried out in 1992. The trial included 1,621 subjects, who were randomised to one of four different treatment groups; daily sunscreen and beta-carotene, daily sunscreen and placebo, usual sunscreen use and beta-carotene, and usual sunscreen use and placebo. The sunscreen intervention was not placebo-controlled for ethical reasons. Firstly, it was thought that an emulsion with no active ingredients may cause increased damage from ultraviolet radiation when the water in the emulsion evaporated, and secondly use of a placebo sunscreen in a population with such high levels of exposure to ultraviolet radiation would lead to sunburn.[298] In this chapter only the sunscreen intervention is considered, and the term *intervention arm* will refer to the arm in which participants were asked to use sunscreen daily, and the term *control arm* will refer to the arm in which participants were asked to use their usual amount of sunscreen.

Follow-up clinics in 1994 and 1996 were used to diagnose new BCCs, and clinically diagnosed BCCs were examined histologically. Participants were also asked at three monthly intervals whether any new BCCs had occurred. Medical records were reviewed for all new BCCs reported in this way. Self-reported BCCs were not included if they were not confirmed by

review of the medical record. Histological reports for self-reported BCCs were obtained where possible. The analysis here includes histologically confirmed BCCs on the head and neck. In order to obtain information on participants' sunscreen use, questionnaires were sent to the participants between August 1996 and March 2002 in approximately six-month intervals. Information on sunscreen use is not available before August 1996.

10.2.2. Analysis method

The purpose of the analyses presented in this chapter is to estimate the effect of sunscreen use on time to first BCC. In Chapter 8, five methods were identified that allowed for departures from randomly allocated treatment in RCTs and survival outcomes.[223, 226, 228, 237, 253] Of these methods, four require a binary definition of compliance, either to estimate a complier average causal effect,[223, 226, 228] or to identify the first time at which a participant is non-compliant.[253]

In the Nambour Trial, questionnaires were used to determine participants' use of sunscreen in approximately six-month intervals between August 1996 and March 2002. They were asked to report how many days per week they were using sunscreen on their face, head or neck. The responses available were never, 1-2 days per week, 3-4 days per week, 5-6 days per week, and every day. Using a method that requires a binary definition of compliance, e.g. defining participants as compliant if they use sunscreen every day, and non-compliant otherwise, does not make full use of the available data. RPSFTMs, described in Section 8.2.3.1 and originally proposed by Robins and Tsiatis,[237] will therefore be used to estimate the effect of sunscreen use on time to first BCC in this chapter.

For this analysis, BCCs occurring within one year of a participant's entry to the study are not included, as these may be due to underlying latent disease at the time of entry. Participants enter the analysis at their date of entry to the study, and are censored at the time of their first BCC more than one year post-randomisation, date of drop-out, or 30th September 2002, whichever comes first.

Table 10.1: Conversion of self-reported sunscreen use into a proportion of time using sunscreen.

Self-reported sunscreen use	Proportion of time using sunscreen
Never	0
1-2 days per week	1.5/7=0.214
3-4 days per week	3.5/7=0.5
5-6 days per week	5.5/7=0.786
Every day	1

The analysis in this chapter will compare sunscreen use with no sunscreen use. In the notation of Equation 8.1, T_{0i} will be the total time not using sunscreen, and T_{1i} will be the total time using sunscreen. These times can be calculated using each subject's self-reported sunscreen use. This requires converting the self-reported sunscreen use to a proportion of time using sunscreen.

The values used for this conversion are shown in Table 10.1.

Missing data on sunscreen use is a large problem in this dataset (see Table 10.5). Three alternative methods are used to deal with this problem, and the results from each compared. The first method will be referred to as *last value carried forward*. For all missing values, other than the response to the first questionnaire in August 1996, the missing value will be replaced by the most recent non-missing value. If information on sunscreen use from the first questionnaire in August 1996 is missing, participants in the intervention arm are assumed to use sunscreen every day, while those in the intervention arm are assumed to never use sunscreen. There was no information on sunscreen use before August 1996, and so it is assumed that throughout this period participants allocated to the intervention arm used sunscreen every day, and those in the control arm never used sunscreen. The reported sunscreen use is assumed to be constant until the date of the next questionnaire. For example, if a participant reports using sunscreen one or two days per week in the August 1996 questionnaire, that subject is assumed to use sunscreen one or two days per week between 31st August 1996 and 31st March 1997, the date of the next questionnaire. Sunscreen use is then converted to a proportion using the values give in Table 10.1, and multiplied by the length of time (in years) between the two questionnaires. This provides an estimate of the total time using sunscreen in the various periods of follow-up, which are then summed to give an estimate of the total time using sunscreen throughout follow-up.

Table 10.2 provides an example of this calculation for subject 10236. Follow-up for this subject started on 15th March 1992 and ended on 14th January 2002, when the subject dropped out of the study. As the subject was randomised to the intervention arm it was assumed that sunscreen was used every day between 15th March 1992 and 31st August 1996. The missing information on sunscreen use for the questionnaires on 31st March 1997 and 30th September 1997 were replaced by the information provided in the August 1996 questionnaire. All other missing values were replaced in a similar way. The information on sunscreen use was converted to a proportion of time using sunscreen, defined by the values in Table 10.1. This proportion was multiplied by the total time of the interval to which it relates to give the total time using sunscreen in that interval (the final column in Table 10.2). Summing this column gives the total time using sunscreen throughout follow-up.

The second method to deal with missing data on sunscreen use will be referred to as *average value post 1996*. For this, the average value of all available responses to the sunscreen use questionnaire is taken and applied to the entire period between August 1996, the date of the first sunscreen questionnaire, and September 2002, the end of the analysis time. For the period between entry to the study and August 1996 people randomised to the intervention arm are assumed to use sunscreen every day, and people randomised to the control arm are assumed to

Table 10.2:Example calculation of total time using (or not using) sunscreen throughout follow-up, using the last value carried forward method for missing values of reported sunscreen use, for subject 10236.

Subject id	Randomised status	Time interval		Analysis time (years)	Reported sunscreen use	Last value carried forward	Proportion of time using sunscreen	Total time using sunscreen (years)
		Start date	End date					
10236	Intervention	15th March 1992	31st August 1996	4.46	-	Every day	1	4.46
		31st August 1996	31st March 1997	0.58	3-4 days	3-4 days	0.50	0.29
		31st March 1997	30th September 1997	0.50	-	3-4 days	0.50	0.25
		30th September 1997	31st March 1998	0.50	-	3-4 days	0.50	0.25
		31st March 1998	30th September 1998	0.50	3-4 days	3-4 days	0.50	0.25
		30th September 1998	31st March 1999	0.50	-	3-4 days	0.50	0.25
		31st March 1999	30th September 1999	0.50	-	3-4 days	0.50	0.25
		30th September 1999	31st March 2000	0.50	Never	Never	0	0
		31st March 2000	30th September 2000	0.50	1-2 days	1-2 days	0.21	0.11
		30th September 2000	31st March 2001	0.50	Never	Never	0	0
		31st March 2001	30th September 2001	0.50	-	Never	0	0
		30th September 2001	14th January 2002	0.29	-	Never	0	0
						Total time using sunscreen throughout follow-up (T_{1i}, in years)		6.11
						Total time not using sunscreen throughout follow-up (T_{0i}, in years)		3.72

never use sunscreen. If there are no responses to the sunscreen use questionnaires, and therefore no values to take an average of, the subject is assumed to use sunscreen every day if they were randomised to the intervention arm, and to never use sunscreen if they were randomised to the control arm. There were only four subjects who had not dropped out or had a BCC before 31st August 1996, and who had no information about sunscreen use.

Table 10.3 illustrates this method with data from subject 10236. For this subject, follow-up started on 15th March 1992 and ended on 14th January 2002 when the subject dropped out of the study. The average value was calculated by summing the proportions of time using sunscreen as defined in Table 10.1 (equal to 1.21 for this subject) and dividing by the number of available responses (equal to five for this subject). This average proportion (0.24) is then applied to all intervals after 31st August 1996, while before 31st August 1996 this subject was assumed to use sunscreen every day because they were randomised to the intervention arm. The total time using sunscreen in each interval was obtained by multiplying the proportion of time using sunscreen by the length of the interval, and the total time using sunscreen throughout follow-up was calculated by summing the times using sunscreen in each interval of follow-up.

These first two methods for dealing with the missing information both make implausible assumptions about sunscreen use between entry to the study and August 1996. It is extremely unlikely that all subjects in the intervention arm were using sunscreen every day, and all subjects in the control arm were never using sunscreen throughout this time. For the third method, therefore, missing information about sunscreen use will be will be dealt with by taking an average of all available responses to the sunscreen questionnaires, and applying this average to the entire analysis time between entry to the study in 1992 and September 2002. This will be referred to as the *average value*. There were 232 people with no information on sunscreen use. For these people, it was assumed that they used sunscreen every day if they were randomised to the intervention arm, and that they never used sunscreen if they were randomised to the control arm. Although this means that sunscreen use is likely to be overestimated in the intervention arm, and underestimated in the control arm for people with no information on sunscreen use, this method may more accurately capture the amount of time that sunscreen was used in both arms of the trial.

Table 10.4 illustrates this method with data from subject 10236. For this subject, follow-up started on 15th March 1992 and ended on 14th January 2002, when the subject dropped out of the study. The average value is calculated by summing the proportions of time using sunscreen (defined in Table 10.1) from the available responses, and dividing by the total number of responses (in this case, five). This average proportion is then applied to all intervals between the start and end of follow-up to provide the amount of time using sunscreen in each interval.

Table 10.3: Example calculation of time using (or not using) sunscreen throughout follow-up, using the average value post 1996 method for missing values of reported sunscreen use, for subject 10236.

Subject id	Randomised status	Time interval		Analysis time (years)	Reported sunscreen use	Proportion of time using sunscreen	Average value post 1996	Total time using sunscreen (years)
		Start date	End date					
10236	Intervention	15th March 1992	31st August 1996	4.46	-		1	4.46
		31st August 1996	31st March 1997	0.58	3-4 days	0.50	0.24	0.14
		31st March 1997	30th September 1997	0.50	-		0.24	0.12
		30th September 1997	31st March 1998	0.50	-		0.24	0.12
		31st March 1998	30th September 1998	0.50	3-4 days	0.50	0.24	0.12
		30th September 1998	31st March 1999	0.50	-		0.24	0.12
		31st March 1999	30th September 1999	0.50	-		0.24	0.12
		30th September 1999	31st March 2000	0.50	Never	0	0.24	0.12
		31st March 2000	30th September 2000	0.50	1-2 days	0.21	0.24	0.12
		30th September 2000	31st March 2001	0.50	Never	0	0.24	0.12
		31st March 2001	30th September 2001	0.50	-		0.24	0.12
		30th September 2001	14th January 2002	0.29	-		0.24	0.07
					Total time using sunscreen throughout follow-up (T_{1i}, in years)			5.75
					Total time not using sunscreen throughout follow-up (T_{0i}, in years)			4.08

Table 10.4: Example calculation of time using (or not using) sunscreen throughout follow-up, using the average value method for missing values of reported sunscreen use, for subject 10236.

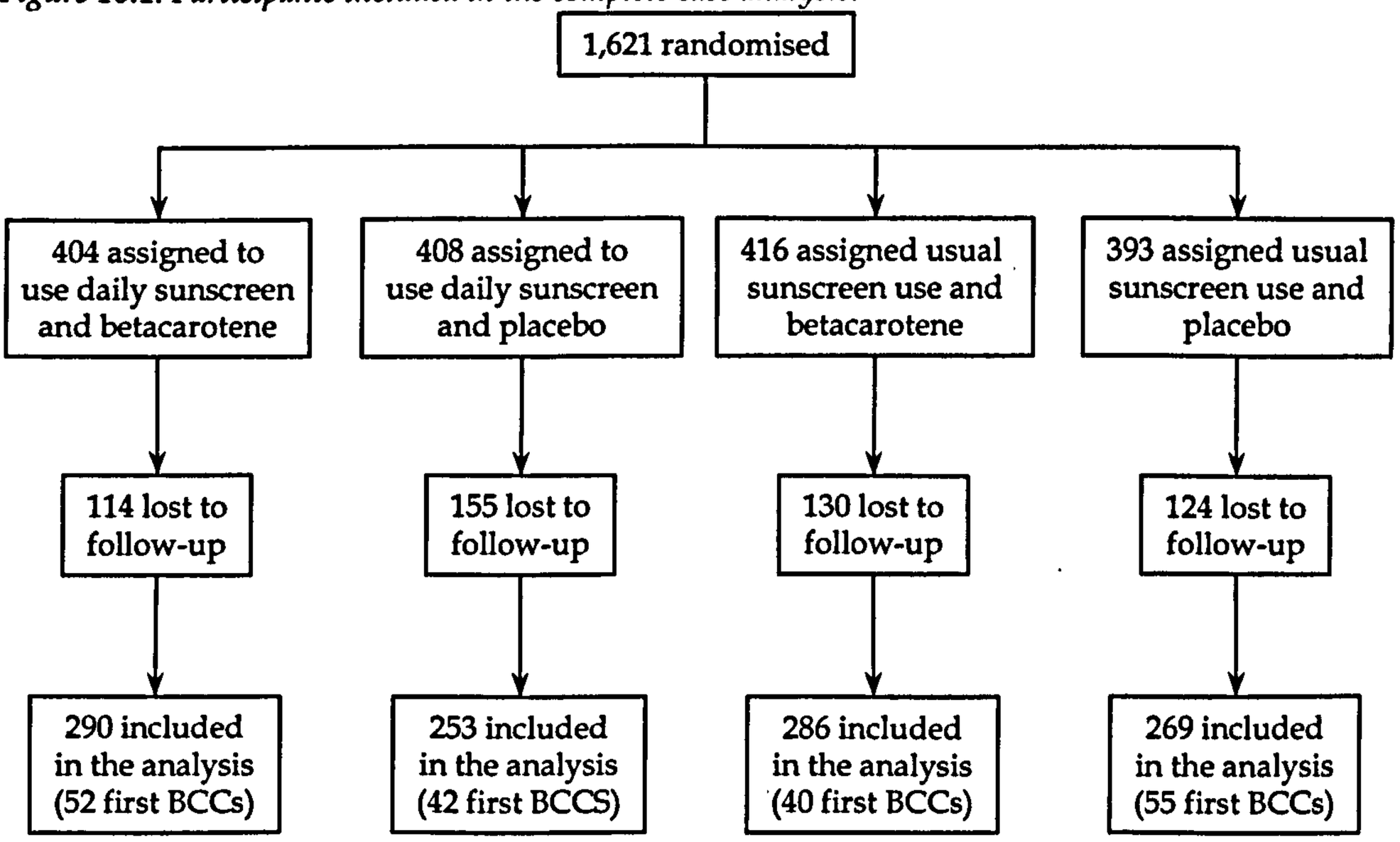
Subject id	Randomised status	Time interval		Analysis time (years)	Reported sunscreen use	Proportion of time using sunscreen	Average value	Total time using sunscreen (years)
		Start date	End date					
10236	Intervention	15th March 1992	31st August 1996	4.46	-		0.24	1.07
		31st August 1996	31st March 1997	0.58	3-4 days	0.50	0.24	0.14
		31st March 1997	30th September 1997	0.50	-		0.24	0.12
		30th September 1997	31st March 1998	0.50	-		0.24	0.12
		31st March 1998	30th September 1998	0.50	3-4 days	0.50	0.24	0.12
		30th September 1998	31st March 1999	0.50	-		0.24	0.12
		31st March 1999	30th September 1999	0.50	-		0.24	0.12
		30th September 1999	31st March 2000	0.50	Never	0	0.24	0.12
		31st March 2000	30th September 2000	0.50	1-2 days	0.21	0.24	0.12
		30th September 2000	31st March 2001	0.50	Never	0	0.24	0.12
		31st March 2001	30th September 2001	0.50	-		0.24	0.12
		30th September 2001	14th January 2002	0.29	-		0.24	0.07
					Total time using sunscreen throughout follow-up (T_{1u} in years)			2.36
					Total time not using sunscreen throughout follow-up (T_{0u} in years)			7.47

Summing these times using sunscreen provides the total time using sunscreen throughout follow-up.

Non-compliance will be used to describe participants in the control arm using sunscreen, and participants in the intervention arm not using sunscreen. The analyses in this chapter will consider non-compliance in both arms simultaneously. It is possible to use RPSFTMs to analyse the effect of non-compliance in each arm separately. This was done, but the results are not provided as they required implausible assumptions about sunscreen use in the randomised arm of the trial in which self-reported sunscreen use was not being used. For example, if non-compliance in the intervention arm only was considered, the subjects in the control arm would be assumed to never use sunscreen, which is extremely implausible.

Two methods to allow for censoring will be used. In the first, censoring will only be due to the administrative end of follow-up, and data for subjects who dropped out before this will not be analysed. This will be called the *complete case analysis*. For the complete case analysis only participants who did not leave the study before 30th September 2002 are included. Between the date of entry into the study and 30th September 2002, 523 participants dropped out, leaving 1,098 subjects for the analysis. Figure 10.1 shows the numbers remaining in each treatment arm, and the number of first BCCs occurring between one year post-randomisation and 30th September 2002.

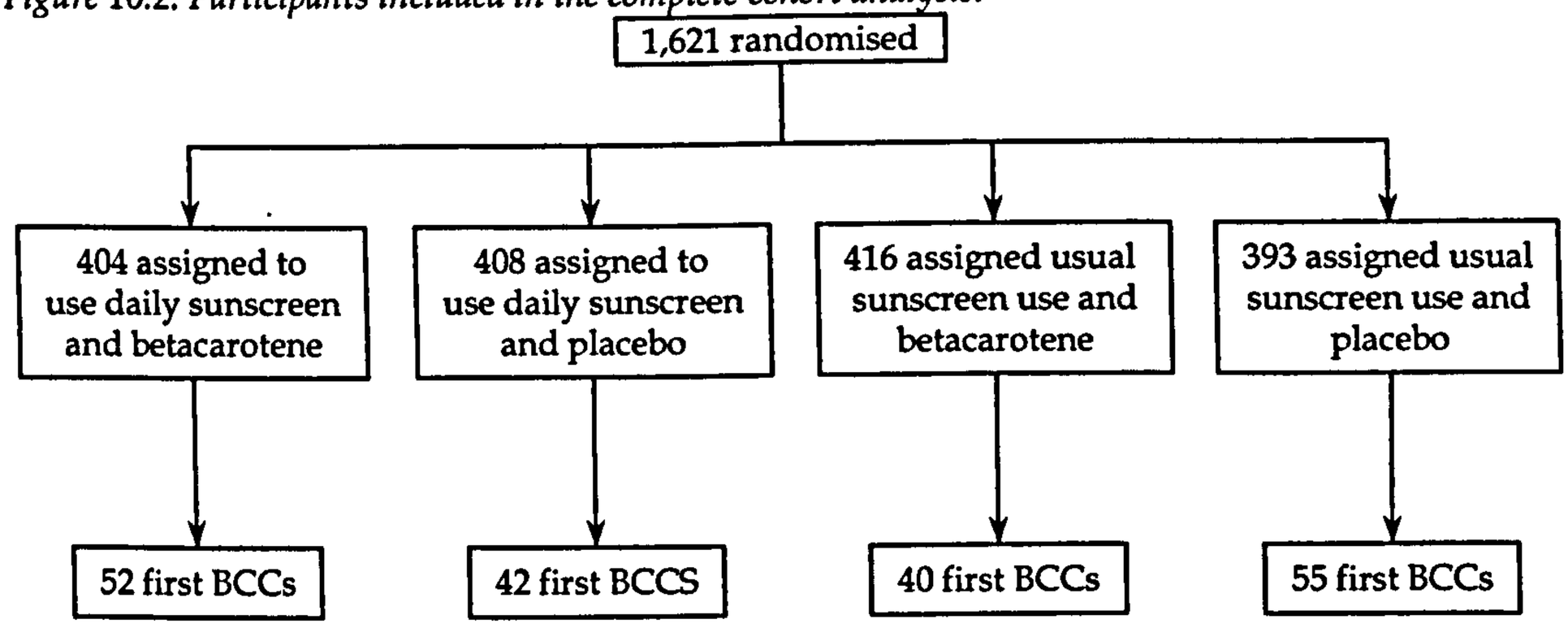
Figure 10.1: Participants included in the complete case analysis.



In the second analysis, it is assumed that loss to follow-up is non-informative and subjects censored before the end of follow-up will be included. This will be called the *complete cohort*

analysis. Figure 10.2 shows the numbers in each treatment arm, and the number of first BCCs occurring between one year post-randomisation and 30th September 2002.

Figure 10.2: Participants included in the complete cohort analysis.



Inverse probability of censoring weights, as discussed in Section 8.2.3.1, are not used for censoring in the analysis in this chapter, as the dataset provided for the analysis contained no information on covariates that predict censoring. A grid search, described in Section 8.2.3.1, was used to obtain the point estimates and 95% CIs for the acceleration parameter from the RPSFTM.

Hazard ratios, and two corresponding 95% confidence intervals (CIs), are calculated for each analysis using the method described in Section 8.2.3.1. The CI based on the p-value of the intention-to-treat (ITT) analysis will be referred to as the *test-based* CI. The second CI, based on 1,000 bootstrapped samples, will be referred to as the *bootstrapped* CI. This 95% CI will be calculated using the bias-corrected and accelerated (BC_a) method,[299] which is a percentile method for estimating CIs from bootstrap samples. To reduce computation time for the bootstrapped 95% CI, $\hat{\beta}$ is estimated using an interval bisection method (described in Section 8.2.3.1), rather than a grid search.

10.2.2.1 The bias-corrected and accelerated method

Using the same notation as Efron and Tibshirani,[299] a bootstrapped (100-2α)% CI is calculated using the BC_a method as $\left(\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}\right)$, where $\hat{\theta}^{*(\alpha)}$ is the αth percentile of the bootstrap replications of the parameter of interest,

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1-\hat{a}\left(\hat{z}_0 + z^{(1-\alpha)}\right)}\right), \text{ and}$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1-\hat{a}\left(\hat{z}_0 + z^{(\alpha)}\right)}\right).$$

In the equations above, $\Phi(x)$ is the standard normal cumulative distribution function, and $z^{(\alpha)}$

is the αth percentile point of a standard normal distribution. The bias-correction, $\hat{z}_0$, is calculated using the formula

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{\theta}^*(b) < \hat{\theta}\}}{B}\right)$$

where $\Phi^{-1}(x)$ is the inverse function of a standard normal cumulative density function, B is the number of bootstrap samples, and the numerator is the number of bootstrap replicates of the parameter of interest that are smaller than the parameter estimate in the original dataset. The acceleration, $\hat{a}$, is calculated using the formula

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right\}^{3/2}}$$

where n is the number of observations in the original dataset, $\hat{\theta}_{(i)}$ is the parameter estimate obtained when the ith observation is deleted from the original dataset, and

$$\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^{n}\hat{\theta}_{(i)}}{n}.$$

10.3. Results

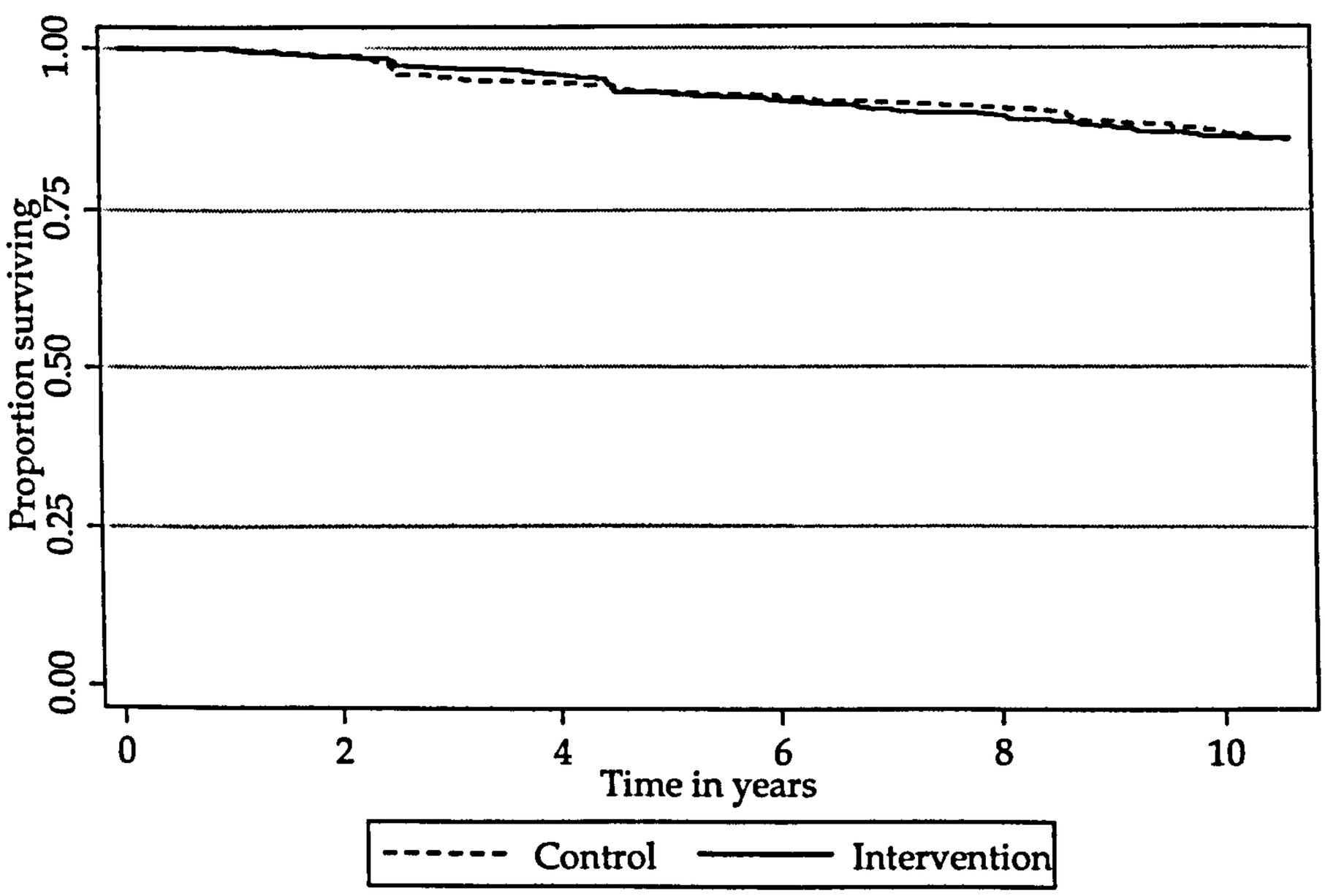
10.3.1. Descriptive analysis

Figure 10.3 shows the Kaplan-Meier survival curves for time to first BCC in the intervention and control groups. There is little difference between the survival curves in the two groups. Using a log-rank test to test for equality of the survivor functions in the two groups gives a p value of 0.98, indicating no evidence of any difference in survivor functions in the two groups. The probability of surviving to ten years without a BCC is estimated to be 0.87 in the control group, and 0.86 in the intervention group.

Because some subjects randomised to the control group used sunscreen, and some subjects randomised to the intervention group did not use sunscreen every day, the survival curves are more similar than they would have been had all subjects in the intervention group used sunscreen every day and all subject in the control group never used sunscreen. A large amount of data is available in the Nambour Trial on the amount of sunscreen used by participants. These data are stratified by randomised group, and summarised in Table 10.5. Considering the data from all subjects simultaneously, responses to the first questionnaire appear to differ from subsequent responses. For example, 7% of participants indicated never using sunscreen in the first questionnaire on 31st August 1996. In the next questionnaire, this proportion increased to 33%. This increase is accompanied by a decrease in the number of missing responses, and a

decrease in the number of people reporting using sunscreen for five days or more per week. Following the first questionnaire, there are no apparent discrepancies between questionnaire responses at any other times. The proportion of missing responses increases over time, as expected due to drop out from the study. The proportion of responses to the other categories in each questionnaire either remains fairly constant, or declines slightly.

Figure 10.3: Kaplan-Meier survival curves for time to first basal cell carcinoma in the control and intervention groups.



Considering reported sunscreen use in the intervention and control arms separately, there is a six-fold increase in the number of participants in the intervention group reporting never using sunscreen between the first and second questionnaires. This is accompanied by an approximately four-fold increase in the number of participants in the control group indicating no sunscreen use, although the absolute increase in the control group is greater than that in the intervention group. The decrease in the overall number of missing responses appears to be mainly due to a decrease in the number of missing responses in the control subjects. Missing responses fell by almost 50% in this group between the first and second questionnaires. The decrease in the number of subjects reporting sunscreen use every day is due to a decrease in the number of intervention subjects reporting this between the first and second questionnaires. In this case, the number reporting every day sunscreen use dropped by just over 50%.

Table 10.5: Responses to sunscreen use questionnaires for the complete cohort, separated by sunscreen allocation. Each cell shows the number of subjects providing a particular response, and the proportion of the responses at each time point.

Sunscreen use	Date of sunscreen use information												
	31/08/96	31/03/97	30/09/97	31/03/98	30/09/98	31/03/99	30/09/99	31/03/00	30/09/00	31/03/01	30/09/01	31/03/02	Total
Control													
Never	87	340	343	360	327	339	321	292	283	275	281	273	3,521
	(0.05)	(0.21)	(0.21)	(0.22)	(0.20)	(0.21)	(0.20)	(0.18)	(0.17)	(0.17)	(0.17)	(0.17)	(0.18)
1-2 days	226	136	123	128	139	121	117	111	109	108	107	101	1,526
	(0.14)	(0.08)	(0.08)	(0.08)	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.08)
3-4 days	60	57	54	41	50	47	46	36	51	43	45	42	572
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
5-6 days	52	34	42	33	47	49	42	41	43	42	31	28	484
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Every day	64	62	57	62	65	62	76	72	62	72	79	90	823
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)	(0.04)
Missing	320	180	190	185	181	191	207	257	261	269	266	275	2,782
	(0.20)	(0.11)	(0.12)	(0.11)	(0.11)	(0.12)	(0.13)	(0.16)	(0.16)	(0.17)	(0.16)	(0.17)	(0.14)
Intervention													
Never	26	193	186	233	202	223	215	198	194	190	190	189	2,239
	(0.02)	(0.12)	(0.11)	(0.14)	(0.12)	(0.14)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
1-2 days	117	182	184	164	170	168	153	140	144	134	130	125	1,811
	(0.07)	(0.11)	(0.11)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)
3-4 days	111	77	75	72	84	60	67	58	60	56	55	54	829
	(0.07)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)
5-6 days	113	61	53	43	47	51	54	38	52	41	42	37	632
	(0.07)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Every day	243	118	116	108	107	104	96	100	84	97	110	103	1,386
	(0.15)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.07)	(0.06)	(0.07)
Missing	202	181	198	192	202	206	227	278	278	294	285	304	2,847
	(0.12)	(0.11)	(0.12)	(0.12)	(0.12)	(0.13)	(0.14)	(0.17)	(0.17)	(0.18)	(0.18)	(0.19)	(0.15)

Rounding may result in proportions that do not sum to one in each column.

10.3.2. Intention-to-treat analysis

The usual method for estimating ITT acceleration factors would be to use a Weibull model. These are not used here as the analysis to estimate the efficacy of sunscreen will use a RPSFTM, and it makes more sense to compare estimated acceleration factors obtained using the same method. To obtain an ITT estimate of the acceleration parameter from a RPSFTM, the possibly counterfactual treatment-free failure times are calculated using Equation 8.2 where, for people in the control arm, T_{0i} is the total analysis time and T_{1i} is zero. Conversely, for people in the intervention arm, T_{1i} is the total analysis time and T_{0i} is zero.

Figure 10.4: Graph of the grid search for the intention-to-treat point estimate and 95% confidence interval for the effect of sunscreen use on time to first basal cell carcinoma.

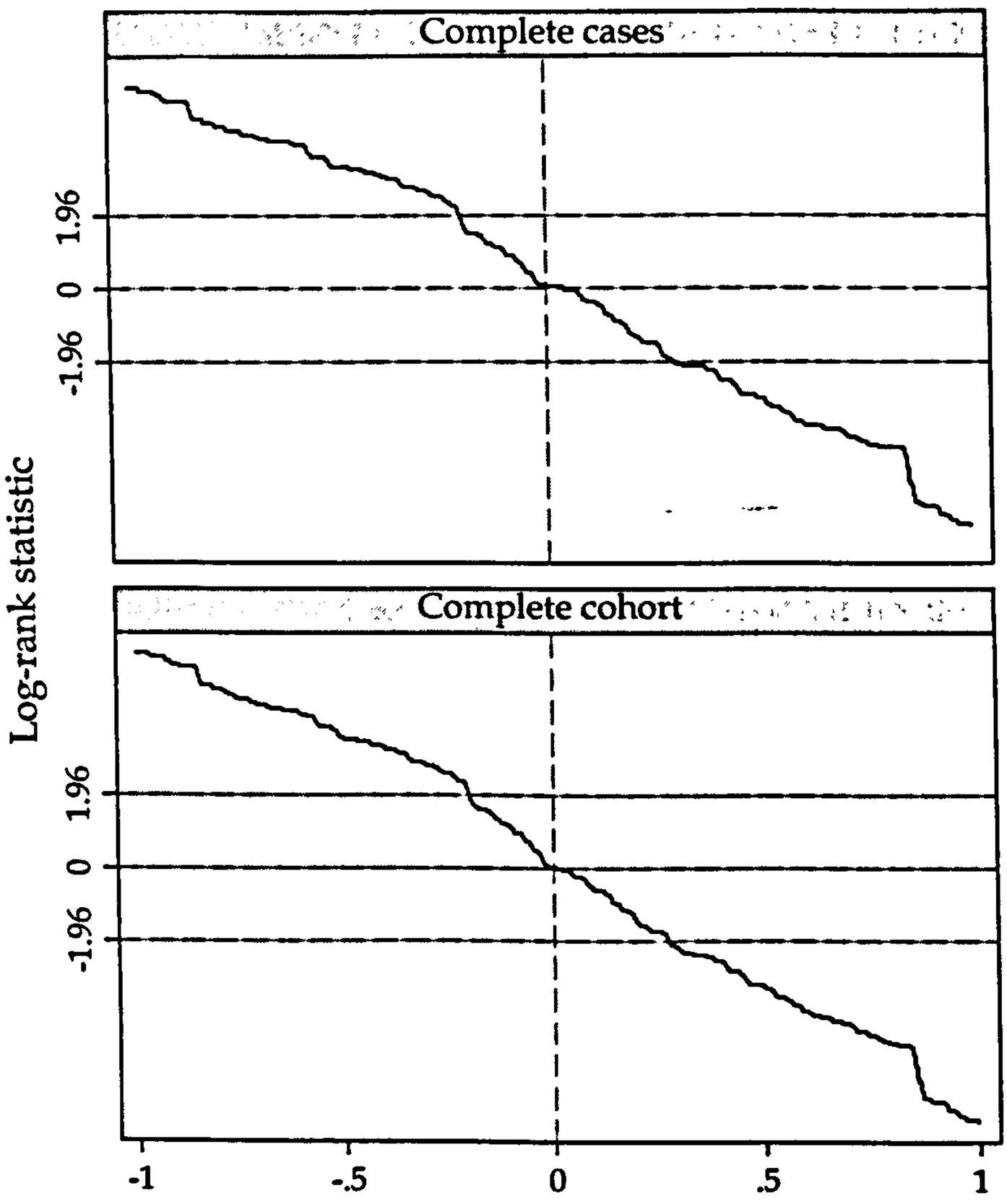


Figure 10.4 shows the results of the grid search to find the ITT point estimate and 95% CI from the RPSFTM. The plotted line shows the value of the log-rank statistic at values of the acceleration parameter between -1 and 1. The horizontal dashed lines on the plot indicate the values at which the log-rank statistic equals 1.96, 0 or -1.96. The point at which the plotted line crosses the value 1.96 defines the lower 95% confidence limit for the acceleration parameter, and

the point at which the plotted line crosses the value -1.96 defines the upper 95% confidence limit for the acceleration parameter. The point at which the plotted line crosses the value 0 provides the point estimate of the acceleration parameter. The vertical dashed line indicates the point at which there is no effect of sunscreen use on time to first BCC. All values of the acceleration parameter to the right of this line indicate a harmful effect of sunscreen use on time to first BCC. Similarly, all values to the left of this vertical line indicate a beneficial effect of sunscreen use on time to first BCC.

The ITT acceleration parameter obtained from the RPSFTM is 0.043 (-0.204-0.297) for the complete cases analysis, and 0.006 (-0.202-0.271) for the complete cohort analysis. The ITT hazard ratio, obtained using a Cox proportional hazards model, is 1.01 (0.76-1.35) for the complete cases analysis, and 1.00 (0.75-1.33) for the complete cohort analysis. The ITT p-value is 0.921 for the complete cases analysis, and 0.983 for the complete cohort analysis.

From these ITT results, and the properties of the RPSFTM, it is possible to predict what results will be obtained when analysing the effect of sunscreen use on time to first BCC. As the ITT estimate will be biased towards the null value as an estimate of treatment efficacy (as subjects in the intervention arm use more sunscreen than subjects in the control arm), allowing for sunscreen use will produce efficacy estimates that are further from the null value than the ITT estimate. In addition, as the RPSFTM preserves the p-value of the ITT analysis, the increase in the efficacy estimate will result in wider 95% CIs with 95% confidence limits that are further from the null value than the ITT 95% confidence limits. As the ITT p-values are large, small increases in the efficacy estimate will result in large increases in the width of the 95% CI.

10.3.3. Complete case analysis

In this section, the effect of sunscreen use on time to first BCC is estimated, using data on people who did not drop out of the trial before 30th September 2002.

Figure 10.5 shows the graphs for the grid searches for the point estimate and 95% CIs for the effect of sunscreen use on time to first basal cell carcinoma for the three different methods of allowing for missing sunscreen use information. Values of the acceleration parameter between -1 and 1 were searched for the last value carried forward and average post 1996 methods. For the average value method, the search was over values of the acceleration parameter between -4 and 4, as upper and lower 95% confidence limits were not found in the original search range. The acceleration parameter point estimates and 95% CIs were derived from these graphs using the method described for the ITT RPSFTM analysis; the point estimate is defined as the value of the acceleration parameter where the log-rank statistic equals zero, and the 95% CI is defined by the values of the acceleration parameter where the log-rank statistic equals 1.96 and -1.96. In contrast with the ITT analysis, in these analyses there may be multiple solutions for the point

estimate and 95% confidence limits. This occurs when the log rank statistic is equal to -1.96, 0 or 1.96 for several different values of the acceleration parameter. These multiple solutions are dealt with in the way described in Section 8.2.3.1. If the log-rank statistic crosses the required values (-1.96, 0 or 1.96) at the points a_0, a_1, ..., a_n, for even n, the estimate is taken to be $\sum_{i=0}^{n}(-1)^i a_i$.

Figure 10.5: Graph of the grid search for the point estimate and 95% confidence interval for the effect of sunscreen use on time to first basal cell carcinoma from a rank preserving structural failure time model, using three methods for missing sunscreen use information.

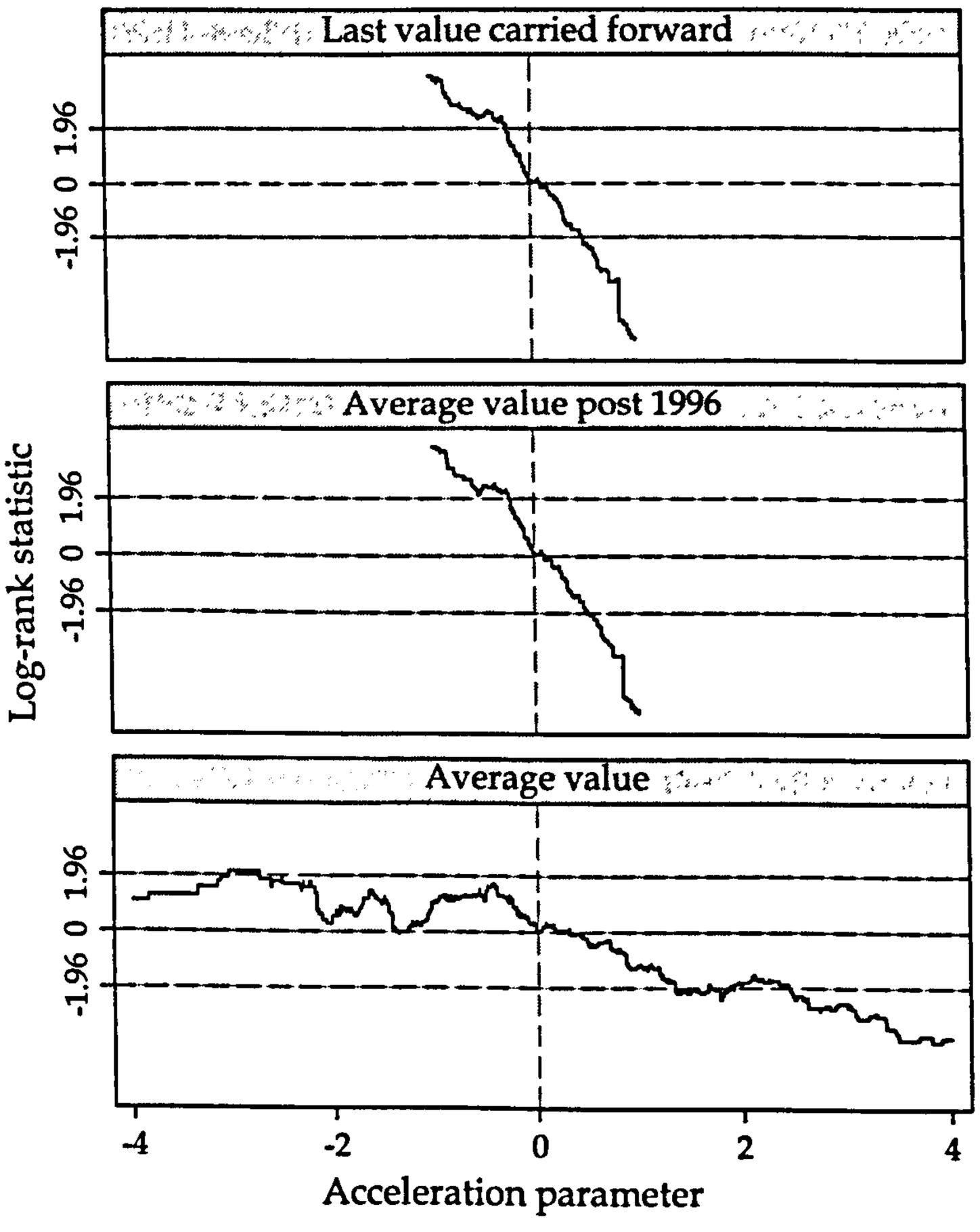
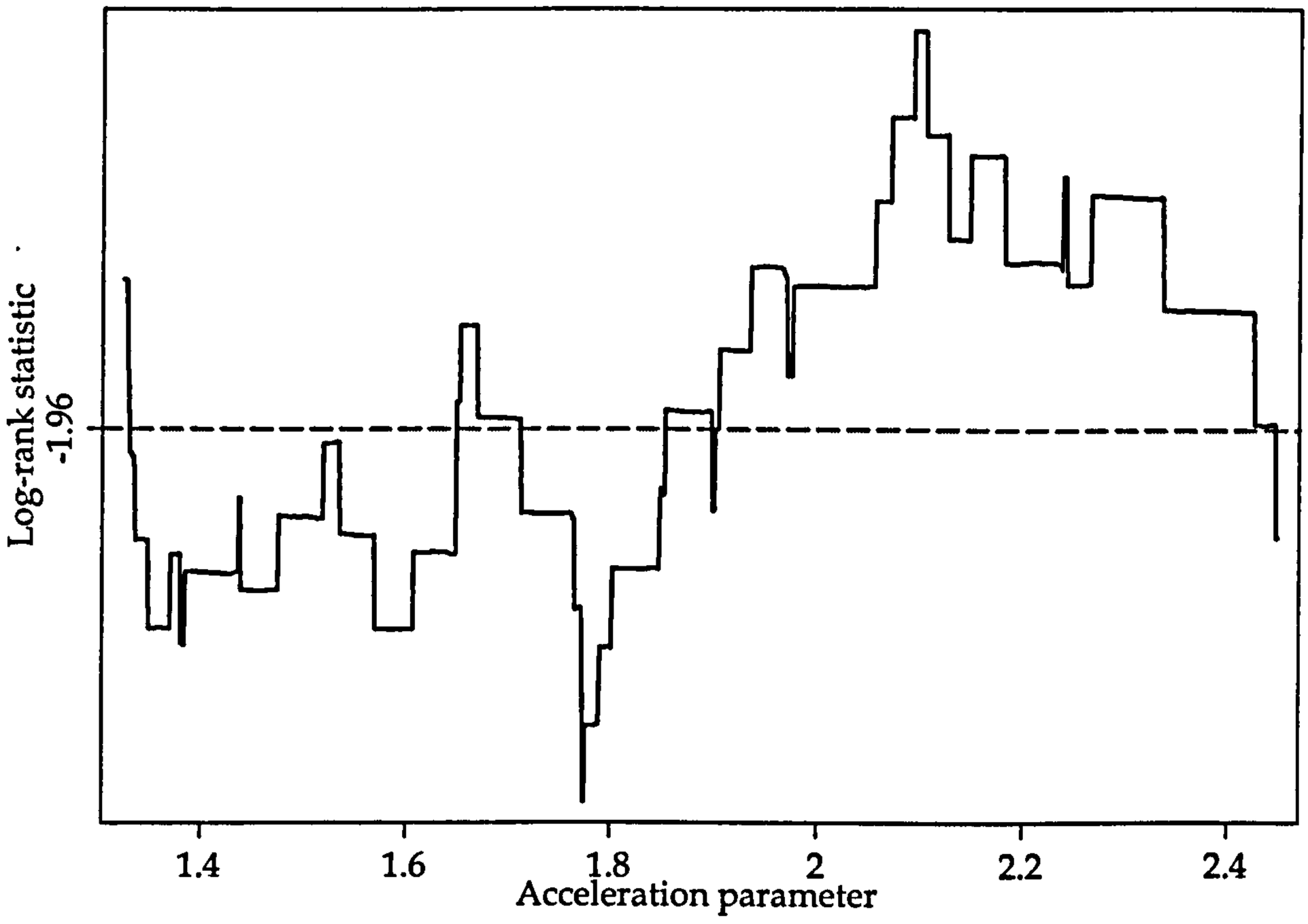


Figure 10.6 shows a magnification of the multiple solutions for the upper 95% confidence limit when the average value method is used to allow for missing information on sunscreen use. The solutions occur at 1.331, 1.650, 1.712, 1.853, 1.898, 1.905 and 2.447. The upper 95% confidence limit for this analysis is therefore defined as 1.98. Multiple solutions also occur for the lower 95% confidence limit for the analysis using the average value method, and for the point estimates for the acceleration parameter for all analyses. The same procedure is used for all multiple solutions occurring throughout this chapter.

Figure 10.6: Magnification of the multiple solutions for the upper 95% confidence limit for the average value method for allowing for missing information on sunscreen use and the complete cases analysis.



Mark and Robins[244] showed that, as compliance with the randomly allocated treatment worsens, the ability of the RPSFTM to distinguish the value of the true acceleration parameter from other values decreases. This is apparent in Figure 10.5 for the average value method. The grid search is non-monotonic, and at no value of the acceleration parameter does the log-rank statistic become greater than, and remain greater than, 1.96. The lower 95% confidence limit for this analysis is therefore defined to be $-\infty$. This means that the data are consistent with sunscreen use being perfectly protective against BCC.

Table 10.6 shows the total amount of time using sunscreen and not using sunscreen in the intervention and control arms for each of the three methods for dealing with missing sunscreen use information. This table shows that the estimated sunscreen use in the control and intervention arms for the last value carried forward and average value from 1996 methods are different. When sunscreen use is estimated using the average value for the entire analysis time, there is very little difference between the estimated sunscreen use in the intervention and control groups. A comparison of the time using sunscreen between the randomised groups will then provide little information about the true value of the acceleration parameter, and leads to the problem of finding the lower 95% confidence limit seen in Figure 10.5.

Table 10.6: Summary of total time using, or not using sunscreen in the intervention and control arms for the three different methods for missing data on sunscreen use.

		Intervention	Control
Last value	Time using sunscreen (years)	3,516	760
	Time not using sunscreen (years)	1,744	4,636
Average value post 1996	Time using sunscreen (years)	3,476	768
	Time not using sunscreen (years)	1,784	4,628
Average value	Time using sunscreen (years)	2,013	1,397
	Time not using sunscreen (years)	3,247	3,999

Table 10.7 shows the effect estimate, acceleration factor and hazard ratio from fitting a RPSFTM to those participants of the Nambour trial who did not drop out before the end of follow-up. The ITT estimate is shown, along with the estimates obtained when actual sunscreen use is considered using the three methods of allowing for missing information on sunscreen use. As predicted in Section 10.3.2, the point estimates allowing for sunscreen use are all further from the null value than the ITT estimate. For the last value carried forward analysis, the point estimate of $\hat{\beta}$ is 2.1 times larger than the ITT point estimate. The hazard ratio also increases from 1.01 from the ITT analysis to 1.07 when sunscreen use is considered. For the average value post 1996 method, the point estimate of $\hat{\beta}$ is 2.8 times larger than the ITT point estimate. The hazard ratio also increases from 1.01 from the ITT analysis to 1.09 when sunscreen use is considered. For the average value method for allowing for missing information on sunscreen use, the point estimate of $\hat{\beta}$ is 6.4 times larger than the ITT point estimate. The hazard ratio also increases from 1.01 from the ITT analysis to 1.32 when sunscreen use is considered.

Table 10.7: Effect estimate, acceleration factor and hazard ratio for the effect of sunscreen use on time to first basal cell carcinoma from a rank preserving structural failure time model, using three methods for missing sunscreen use information.

Analysis	Effect estimate $\hat{\beta}$	Acceleration factor	Hazard ratio	
	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)[a]	(95% CI)[b]
ITT	0.043 (-0.204-0.297)	1.04 (0.82-1.35)	1.01 (0.76-1.35)	
Last value carried forward	0.091 (-0.230-0.503)	1.10 (0.79-1.65)	1.07 (0.29-3.88)	(0.75-1.74)
Average value post 1996	0.119 (-0.234-0.515)	1.13 (0.79-1.67)	1.09 (0.20-6.04)	(0.75-1.76)
Average value	0.277 (-∞-1.980)	1.32 (0-7.24)	1.32 (0.01-306.04)	(0-5.49)*

[a] *Test-based 95% confidence interval, with the same p-value as the ITT estimate.*
[b] *Bootstrapped 95% confidence interval, based on 1,000 bootstrap samples.*
* *Based on 985 bootstrap samples.*

All 95% CIs for the analyses allowing for sunscreen use are wider than the 95% CI for the ITT analysis. This was also predicted in Section 10.3.2. The test-based 95% CIs for the hazard ratios when sunscreen use is considered are very wide. Because the p-value for the ITT estimate is large (p=0.921), small increases in the hazard ratio result in a large increase in the width of the

test-based 95% CI. This is particularly evident when the average value method is used to allow for missing information on sunscreen use. The test-based 95% CI for the hazard ratio provides almost no information in this situation. The bootstrapped 95% CI is generally narrower than the test-based 95% CI and may therefore be preferable to the test-based 95% CI, as it is not influenced by large p-values. When deriving the bootstrapped 95% CI for the hazard ratio for the average value method, there were 15 bootstrapped datasets for which an estimate for the acceleration parameter could not be found. This bootstrapped 95% CI is therefore based on only 985 samples. The CIs will not be discussed further in this chapter.

Table 10.8: The effects of recensoring in the complete cases dataset, using three methods for missing sunscreen use information.

Analysis		Control	Intervention
Last value carried forward	Number of observations recensored	348	452
	Total time	5,468.48	5,595.40
	Total time lost	69.16	294.80
	Events lost	2	3
Average value post 1996	Number of observations recensored	348	453
	Total time	5,493.11	5,699.71
	Total time lost	92.49	387.59
	Events lost	2	4
Average value	Number of observations recensored	353	428
	Total time	5,842.07	5,902.99
	Total time lost	414.30	581.41
	Events lost	7	7

One of the main issues with analysis of this type is that recensoring is required to deal with informative censoring. This results in a loss of information, and the estimated acceleration parameters and 95% CIs displayed in Figure 10.5 and Table 10.7 may be due to recensoring rather than a true effect of sunscreen use on time to BCC. Table 10.8, therefore, shows the number of observations that are recensored for the analysis of sunscreen use, the total analysis time that would have been used had there been no recensoring, the analysis time lost due to recensoring, and the number of events lost due to recensoring for the three methods for allowing for missing information on sunscreen use. For the last value carried forward method 800 (72.9%) observations were recensored. This, however, only resulted in the loss of 3.3% of the analysis time. Only five (2.6%) of the 189 first BCCs were lost due to recensoring. For the average value post 1996 method 801 (73.0%) observations were recensored. This, however, only resulted in the loss of 4.3% of the analysis time. Only six (3.2%) of the 189 first BCCs were lost due to recensoring. It would seem that the results presented in Table 10.7 and Figure 10.5 are unlikely to be greatly affected by recensoring for these two analysis methods. For the average value method, however, 781 (71.1%) of the observations were recensored. This resulted in the loss of 8.5% of the analysis time. Fourteen (7.4%) of the 189 first BCCs were lost due to recensoring. It would seem that the results presented in Table 10.7 and Figure 10.5 are more affected by recensoring than the results described for the other two methods for dealing with missing information on sunscreen use.

10.3.4. Complete cohort analysis

In this section, the effect of sunscreen use on time to first BCC is estimated, using data from all 1,621 participants in the Nambour Trial. As the estimated acceleration parameters obtained in these analyses are similar to those obtained in Section 10.3.3, the results will be presented more concisely.

Figure 10.7: Graph of the grid search for the point estimate and 95% confidence interval for the effect of sunscreen use on time to first basal cell carcinoma from a rank preserving structural failure time model, using three methods for missing sunscreen use information.

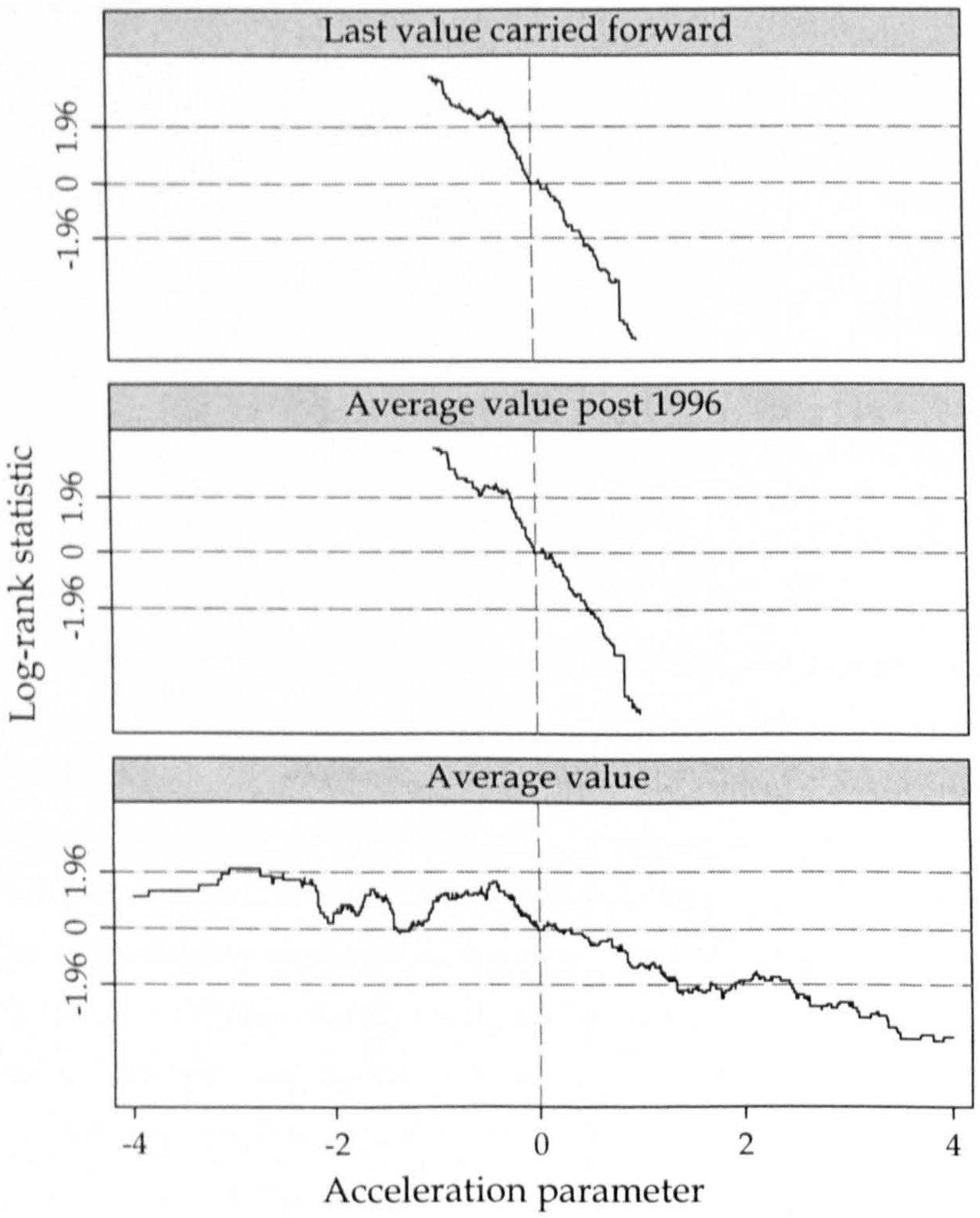


Figure 10.7 shows the graphs for the grid searches for the point estimate and 95% CIs for the effect of sunscreen use on time to first basal cell carcinoma for the three different methods of allowing for missing sunscreen use information. The acceleration parameter point estimates and 95% CIs were derived from these graphs using the method described for the ITT RPSFTM analysis. There are multiple solutions for the point estimate of the acceleration parameter for all three analyses. There are multiple solutions for the upper 95% confidence limit for the average value post 1996 and the average value analyses. For the last value carried forward and the average value analyses, there are multiple solutions for the lower 95% confidence limit. These

multiple solutions are dealt with in the way described in Section 8.2.3.1 and Section 10.3.3.

The inability of the RPSFTM to distinguish the value of the true acceleration parameter from other values is again apparent in Figure 10.7 for the average value method. The grid search is non-monotonic, and at no value of the acceleration parameter does the log rank statistic become greater than, and remain greater than, 1.96. The lower 95% confidence limit for this analysis is therefore defined to be $-\infty$. Table 10.9 shows that there is very little difference between the estimated sunscreen use in the intervention and control groups for the average value method, which leads to the problem of finding the lower 95% confidence limit for this analysis. For the other two analyses, the estimated time using sunscreen is different between the intervention and control groups, and the same problems do not occur.

Table 10.9: Summary of total time using, or not using sunscreen in the intervention and control arms for the three different methods for missing data on sunscreen use.

		Intervention	Control
Last value	Time using sunscreen (years)	4,559	875
	Time not using sunscreen (years)	2,139	5,908
Average value post 1996	Time using sunscreen (years)	4,515	883
	Time not using sunscreen (years)	2,182	5,899
Average value	Time using sunscreen (years)	2,644	1,639
	Time not using sunscreen (years)	4,054	5,144

Table 10.10 shows the effect estimate, acceleration factor and hazard ratio from fitting a RPSFTM to all participants of the Nambour trial. The ITT estimate is shown, along with the estimates obtained when actual sunscreen use is considered using the three methods of allowing for missing information on sunscreen use. For the last value carried forward analysis, the point estimate of $\hat{\beta}$ is 12.2 times larger than the ITT point estimate. The hazard ratio also increases from 1.00 from the ITT analysis to 1.05 when sunscreen use is considered. For the average value post 1996 method, the point estimate of $\hat{\beta}$ is 12.7 times larger than the ITT point estimate. The hazard ratio also increases from 1.00 from the ITT analysis to 1.05 when sunscreen use is considered. Although the relative increases in the estimated acceleration factors are much greater for the complete cohort analysis than for the complete cases analysis, the estimated acceleration factors obtained from the two methods are similar. For the average value method for allowing for missing information on sunscreen use, the point estimate of $\hat{\beta}$ is 9.0 times larger than the ITT point estimate. The hazard ratio also increases from 1.00 from the ITT analysis to 1.07 when sunscreen use is considered. These are very different estimates from those obtained in the complete cases analyses, but it has already been seen that the model does not perform well when using the average sunscreen use over the entire analysis period, and therefore little should be inferred from the difference between the estimated acceleration parameters from the two analyses.

Table 10.10: Effect estimate, acceleration factor and hazard ratio for the effect of sunscreen use on time to first basal cell carcinoma from a rank preserving structural failure time model, using three methods for missing sunscreen use information.

Analysis	Effect estimate $\hat{\beta}$	Acceleration factor	Hazard ratio	
	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)[a]	(95% CI)[b]
ITT	0.006 (-0.202-0.271)	1.01 (0.82-1.31)	1.00 (0.75-1.33)	
Last value carried forward	0.073 (-0.244-0.503)	1.08 (0.78-1.65)	1.05 (0.01-85.66)	(0.77-1.82)
Average value post 1996	0.076 (-0.245-0.510)	1.08 (0.78-1.67)	1.05 (0.01-124.49)	(0.76-1.84)
Average value	0.054 (-∞-1.872)	1.06 (0-6.50)	1.07 (0.00-704.09)	(0-4.26)*

[a] *Test-based 95% confidence interval, with the same p-value as the ITT estimate.*
[b] *Bootstrapped 95% confidence interval, based on 1,000 bootstrap samples.*
* *Based on 986 bootstrap samples*

Table 10.11 shows the number of observations that are recensored for the analysis of sunscreen use, the total analysis time that would have been used had there been no recensoring, the analysis time lost due to recensoring, and the number of events lost due to recensoring for the three methods for allowing for missing information on sunscreen use. For the last value carried forward method 1,138 (70.2%) observations were recensored. This, however, only resulted in the loss of 2.7% of the analysis time. Only three (1.6%) of the 189 first BCCs were lost due to recensoring. For the average value post 1996 method 1,139 (70.3%) observations were recensored. This, however, only resulted in the loss of 2.8% of the analysis time. Only three (1.6%) of the 189 first BCCs were lost due to recensoring. For the average value method 1,112 (68.6%) observations were recensored. This, however, only resulted in the loss of 1.6% of the analysis time. Only two (1.1%) of the 189 first BCCs were lost due to recensoring. It would seem that the results presented in Table 10.10 and Figure 10.7 are unlikely to be greatly affected by recensoring for these analyses.

Table 10.11: The effects of recensoring in the complete cohort, using three methods for missing sunscreen use information.

Analysis		Control	Intervention
Last value carried forward	Number of observations recensored	419	719
	Total time	6,849.08	7,042.99
	Total time lost	63.55	313.06
	Events lost	2	1
Average value post 1996	Number of observations recensored	420	719
	Total time	6,852.60	7,054.30
	Total time lost	66.71	323.50
	Events lost	2	1
Average value	Number of observations recensored	439	673
	Total time	6,873.78	6,844.47
	Total time lost	84.03	134.77
	Events lost	2	0

10.4. Discussion

10.4.1. Summary of results

For the analyses on both the complete cases and the complete cohort, there was no evidence for an effect of sunscreen on time to first BCC from the ITT analysis. The ITT point estimates were slightly greater than the null value, and therefore estimating the efficacy of sunscreen use on time to BCC using a RPSFTM produced point estimates that were further from the null than the ITT point estimate. As RPSFTMs respect the randomisation and preserve the ITT p-value, the increases in the point estimates were accompanied by an increase in the width of the 95% CI when compared to the ITT 95% CI. The additional amount of non-compliance created by using an average of the available information on sunscreen use, and applying this average to the whole follow-up period, meant that the model was unable to distinguish the true value of the acceleration parameter from any other values.[244] This resulted in infinite lower 95% confidence limits.

Recensoring is required to avoid biased estimates from RPSFTMs due to informative censoring. This, however, results in loss of information and may affect effect estimates. For most of the analyses presented in this chapter, loss of information was small and it is unlikely that the effect estimates were greatly affected by this. For the analysis on the complete cases, using an average value of sunscreen use over the entire follow-up period to estimate the total time using sunscreen, 8.4% of the analysis time and 7.4% of the BCCs were lost due to recensoring. This may have influenced the results obtained from the RPSFTM, and account for the large difference between the ITT estimate and the estimate obtained from the RPSFTM.

10.4.2. Strengths and weaknesses

The major strength of methods of analysing departures from randomised exposure in RCTs is that the efficacy of treatment can be estimated, rather than the effectiveness of the randomisation itself. This strength is, however, restricted by the data. If an ITT analysis provides no evidence of an effect of treatment on outcome, neither will a compliance analysis. In the analyses presented in this chapter, there was no evidence of a treatment effect for the ITT analyses, and the ITT point estimate was very close to the null value for both analyses presented here. The main use of a RPSFTM in this situation is to investigate the effect of departures from randomised exposure on the 95% CI, and to find a range of parameter values that more truly reflects the efficacy of exposure.

Data on use of sunscreen was provided by participant self-report. This is likely to be subject to some misclassification, and will therefore not be an accurate measurement of the amount of sunscreen used. The effects of exposure misclassification on parameter estimates for standard

regression models have been well discussed (see Section 2.3.2), but similar investigations have not been carried out for RPSFTMs. It may be the case that exposure misclassification has little or no effect on parameter estimates obtained from RPSFTMs, a result that was demonstrated by Goetghebeur and Vansteelandt[256] in the context of structural nested mean models.

There was a large amount of missing information on sunscreen use in this dataset, due to both missing responses to the questionnaires, and the fact that information on sunscreen did not start to be collected until August 1996. Three methods were used to estimate the total time using sunscreen allowing for this missing data. In the first method, the most recently reported amount of sunscreen use was used. The second and third methods calculated an average of all available sunscreen use information, and applied this to the period between 31st August 1996 and 30th September 2002, and for the entire follow-up time respectively. For the first two methods, participants randomised to the intervention arm were assumed to use sunscreen every day, while those randomised to the control arm were assumed to never use sunscreen between the date of entry to the study and 31st August 1996. This is unlikely to have occurred in reality. Participants randomised to the control arm were told to use their usual amount of sunscreen. It is therefore likely that participants in the control arm were using sunscreen, at least for part of the time between their entry to the study and August 1996. The third method therefore seems the most plausible way to estimate the time using sunscreen, but the similarity between the amount of sunscreen used in the intervention and control arms meant that the model was unable to distinguish the true value of the acceleration parameter from other values. There is also likely to be some inaccuracy in the estimated amount of time using sunscreen obtained from this third method, as sunscreen use habits may have changed over the course of the trial, and the average of the reported sunscreen use after 1996 may not accurately reflect the amount of time that sunscreen was used in the first part of the trial. Due to the difficulty of accurately estimating the amount of time using sunscreen, it seems unlikely that the true effect of sunscreen use on time to BCC has been estimated.

To use the self-reported sunscreen use in the analyses presented here, they were converted into proportions (as shown in Table 10.1). Different proportions could have been calculated for each category of sunscreen use. For example, using sunscreen one or two days per week was converted into a proportion equal to $1.5/7$. This proportion could instead have been calculated as $1/7$ or $2/7$. Changing the way the proportion was calculated is unlikely to make a substantial difference to the results presented in this chapter.

For the complete cohort analysis (presented in Section 10.3.4), drop out was assumed to be non-informative. If this assumption is not true the results presented in Section 10.3.4 will be biased, as censoring prior to the end of follow up will have been dealt with incorrectly. A different method is required to deal with informative censoring.

10.4.3. Future research

Further research could be carried out on data from the Nambour Trial. A sensitivity analysis could be performed to investigate the impact of misclassification of self-reported sunscreen use on the RPSFTM results presented in this chapter. Misclassification in repeated measures of polytomous variables could be investigated using an adaptation of the method proposed by Lash and Fink.[61]

The RPSFTM, proposed by Robins and Tsiatis,[237] has been extended to include time to repeated outcome events by Vandebosch, Goetghebeur and Van Damme[250] and Matsui.[249] The Nambour Trial data contain information on all BCC occurrences during the trial. These methods for analysing compliance and time to repeated outcomes could be used.

Data on other skin cancers, such as malignant melanoma and SCC, are available from the Nambour Trial. Similar analyses to those presented in this chapter could be carried out to investigate the effect of sunscreen use on these cancers.

BLANK IN ORIGINAL

Chapter 11.

Summary

This thesis has examined three aspects of causal modelling in epidemiology; the effects of measurement error in continuous exposures and confounders on estimated exposure-outcome odds ratios (Part A), correcting exposure effect estimates for the effects of measurement error in continuous exposures and confounders (Part B), and estimating the efficacy of treatment in randomised trials in which there are departures from the randomly allocated treatment (Part C). Although these have been treated as distinct aspects within the topic of causal modelling, there are similar points to consider for all three topics, particularly in relation to the implications for study design and analysis (described below). These three topics are not, however, a comprehensive overview of all aspects of causal modelling in epidemiology. For example, estimating exposure effects in the presence of time-dependent confounding is one area that has not been considered.

Some wider implications for study design and analysis have been highlighted in this thesis. Sensitivity analysis is not the best way to allow for measurement error in observational studies. The possibility of measurement error should be considered, for all variables, at the beginning of the study and steps should be taken to minimise it as far as possible. Studies should be designed to allow the use of methods that correct for measurement error (see Chapter 4), for example via internal or external validation of measurements, or via replicate measurements. Study design is also an important consideration for randomised trials in which there may be departures from the randomly allocated treatment. The design should then include assessment, as accurately as possible, of the amount of treatment received by each trial participant.

In order to correct for measurement error, it is important to consider what is the causal exposure variable of interest. For example, in Chapter 7, the long-term average of the variables measured with error was considered to be the causal exposure. In this context, the error variance will include biological variation as well as any errors in measurement, and replicate measurements should be made at sufficiently large intervals to incorporate such variation. If the causal variable is the amount of exposure on a particular day, biological variation will not be included in the estimate of the measurement error variance, and replicate measurements should be made on the same day. Similar considerations apply to analyses of treatment efficacy in the presence of departures from randomly allocated treatment. In order to estimate the causal effect of treatment, the causal treatment (for example, the dose of drug consumed in a given time period) should be clearly defined, and the trial designed in such a way that the causal treatment is measured as precisely as possible.

References

1. Beaglehole, R., Bonita, R., and Kjellstrom, T., (1993). Causation in epidemiology. In: *Basic Epidemiology*, Chapter 5, World Health Organisation, Geneva.

2. Kelsey, J. L., Petitti, D. B., and King, A. C., (1998). Key methodologic concepts and issues. In: *Applied Epidemiology*, Chapter 2, R. C. Brownson and D. B. Petitti (editors), Oxford University Press, New York.

3. Greenland, S. and Robins, J. M., (1986). *Identifiability, exchangeability, and epidemiologic confounding*. International Journal of Epidemiology 15: 413-419.

4. Jurek, A. M., Greenland, S., Maldonado, G. et al., (2005). *Proper interpretation of non-differential misclassification effects: expectations vs observations*. International Journal of Epidemiology 34: 680-687.

5. Khaw, K. T., Bingham, S., Welch, A. et al., (2001). *Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study*. Lancet 357: 657-663.

6. Rimm, E. B., Stampfer, M. J., Ascherio, A. et al., (1993). *Vitamin E consumption and the risk of coronary heart disease in men*. New England Journal of Medicine 328: 1450-1456.

7. Stampfer, M. J., Hennekens, C. H., Manson, J. E. et al., (1993). *Vitamin E consumption and the risk of coronary disease in women*. New England Journal of Medicine 328: 1444-1449.

8. Heart Protection Study Collaborative Group, (2002). *MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20536 high-risk individuals: a randomised placebo-controlled trial*. Lancet 360: 23-33.

9. Collaborative Group of the Primary Prevention Project, (2001). *Low-dose aspirin and vitamin E in people at cardiovascular risk: a randomised trial in general practice*. Lancet 357: 89-95.

10. Davey Smith, G., (2001). *Reflections on the limitations to epidemiology*. Journal of Clinical Epidemiology 54: 325-331.

11. Lawlor, D. A., Davey Smith, G., Bruckdorfer, K. R. et al., (2004). *Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence?* Lancet 363: 1724-1727.

12. Eidelman, R. S., Hollar, D., Hebert, P. R. et al., (2004). *Randomized trials of vitamin E in the treatment and prevention of cardiovascular disease*. Archives of Internal Medicine 164: 1552-1556.

13. Morabia, A., (1995). *Poppers, Kaposi's sarcoma, and HIV infection: empirical example of a strong confounding effect?* Preventive Medicine 24: 90-95.

14. Khaw, K. T., Day, N., Bingham, S. et al., (2004). *Observational versus randomised trial evidence*. Lancet 364: 753-754.

15. Bross, I., (1954). *Misclassification in 2x2 tables*. Biometrics 10: 478-486.

16. Newell, D. J., (1963). *Misclassification in 2x2 tables*. Biometrics 19: 187-188.

17. Phillips, A. N. and Davey Smith, G., (1993). *The design of prospective epidemiological studies:*

More subjects or better measurements? Journal of Clinical Epidemiology 46: 1203-1211.

18. Greenland, S., (1980). *The effect of misclassification in the presence of covariates.* American Journal of Epidemiology 112: 564-569.

19. Flegal, K. M., Brownie, C., and Haas, J. D., (1986). *The effects of exposure misclassification on estimates of relative risk.* American Journal of Epidemiology 123: 736-751.

20. Marshall, J. R. and Hastrup, J. L., (1996). *Mismeasurement and the resonance of strong confounders: uncorrelated errors.* American Journal of Epidemiology 143: 1069-1078.

21. Davidov, O., Faraggi, D., and Reiser, B., (2003). *Misclassification in logistic regression with discrete covariates.* Biometrical Journal 45: 541-553.

22. Sorahan, T. and Gilthorpe, M. S., (1994). *Nondifferential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion.* Occupational and Environmental Medicine 51: 839-840.

23. Fung, K. Y. and Howe, G. R., (1984). *Methodological issues in case-control studies III:- The effect of joint misclassification of risk factors and confounding factors upon estimation and power.* International Journal of Epidemiology 13: 366-370.

24. Dosemeci, M., Wacholder, S., and Lubin, J. H., (1990). *Does nondifferential misclassification of exposure always bias a true effect toward the null value.* American Journal of Epidemiology 132: 746-748.

25. Gilbert E.S., (1991). *Re: "Does nondifferential misclassification of exposure always bias a true effect toward the null value?" (Letter).* American Journal of Epidemiology 134: 440-441.

26. Weinberg, C. R., Umbach, D. M., and Greenland, S., (1994). *When will nondifferential misclassification of an exposure preserve the direction of a trend?* American Journal of Epidemiology 140: 565-571.

27. Brenner, H., (1992). *Notes on the assessment of trend in the presence of nondifferential exposure misclassification.* Epidemiology 3: 420-427.

28. Wacholder, S., Dosemeci, M., and Lubin, J. H., (1991). *Blind assignment of exposure does not always prevent differential misclassification.* American Journal of Epidemiology 134: 433-437.

29. Flegal, K. M., Keyl, P. M., and Nieto, F. J., (1991). *Differential misclassification arising from nondifferential errors in exposure measurement.* American Journal of Epidemiology 134: 1233-1244.

30. Gustafson, P. and Le, N. D., (2002). *Comparing the effects of continuous and discrete covariate mismeasurement, with emphasis on the dichotomization of mismeasured predictors.* Biometrics 58: 878-887.

31. Spearman, C., (1904). *The proof and measurement of association between two things.* American Journal of Psychology 15: 72-101.

32. Armstrong, B. G., Whittemore, A. S., and Howe, G. R., (1989). *Analysis of case-control data with covariate measurement error: application to diet and colon cancer.* Statistics in Medicine 8: 1151-1163.

33. Armstrong, B. G., (1990). *The effects of measurement errors on relative risk regressions.* American Journal of Epidemiology 132: 1176-1184.

34. Clayton, D. and Gill, C., (1991). Covariate measurement errors in nutritional epidemiology: effects and remedies. In: *Design Concepts in Nutritional Epidemiology,*

Chapter 3, B. M. Margetts and M. Nelson (editors), Oxford University Press, Oxford.

35. Wacholder, S., (1995). *When measurement errors correlate with truth: surprising effects of nondifferential misclassification*. Epidemiology 6: 157-161.

36. Marshall, J. R., Hastrup, J. L., and Ross, J. S., (1999). *Mismeasurement and the resonance of strong confounders: correlated errors*. American Journal of Epidemiology 150: 88-96.

37. Kipnis, V., Freedman, L. S., Brown, C. C. et al., (1997). *Effect of measurement error on energy-adjustment models in nutritional epidemiology*. American Journal of Epidemiology 146: 842-855.

38. Walker, A. M. and Lanes, S. F., (1991). *Misclassification of covariates*. Statistics in Medicine 10: 1181-1196.

39. Becher, H., (1992). *The concept of residual confounding in regression-models and some applications*. Statistics in Medicine 11: 1747-1758.

40. Brenner, H., (1993). *Bias due to nondifferential misclassification of polytomous confounders*. Journal of Clinical Epidemiology 46: 57-63.

41. Phillips, A. N. and Davey Smith, G., (1992). *Bias in relative odds estimation owing to imprecise measurement of correlated exposures*. Statistics in Medicine 11: 953-961.

42. Snedecor, G. W. and Cochran, W. G., (1980). Multiple linear regression. In: *Statistical Methods*, 7th edition, Chapter 17, The Iowa State University Press, Ames, Iowa, U.S.A.

43. Fewell, Z., Davey Smith, G., and Sterne, J. A. C., (in press). *The impact of residual and unmeasured confounding in epidemiological studies; a simulation study*. American Journal of Epidemiology.

44. Kirkwood, B. R. and Sterne, J. A. C., (2003). Measurement error: assessment and implications. In: *Essential Medical Statistics*, 2nd edition, Chapter 36, Blackwell Science Ltd., Oxford.

45. Satia-Abouta, J., Patterson, R. E., King, I. B. et al., (2003). *Reliability and validity of self-report of vitamin and mineral supplement use in the vitamins and lifestyle study*. American Journal of Epidemiology 157: 944-954.

46. Schroder, H., Covas, M. I., Marrugat, J. et al., (2001). *Use of a three-day estimated food record, a 72-hour recall and a food-frequency questionnaire for dietary assessment in a Mediterranean Spanish population*. Clinical Nutrition 20: 429-437.

47. Friesema, I. H. M., Veenstra, M. Y., Zwietering, P. J. et al., (2004). *Measurement of lifetime alcohol intake: Utility of a self- administered questionnaire*. American Journal of Epidemiology 159: 809-817.

48. Chinn, S. and Schouten, J. P., (2005). *Reproducibility of non-specific bronchial challenge in adults: implications for design, analysis and interpretation of clinical and epidemiological studies*. Thorax 60: 395-400.

49. Freeman, R., Chase, K. P., and Risk, M. R., (2003). *Quantitative sensory testing cannot differentiate simulated sensory loss from sensory neuropathy*. Neurology 60: 465-470.

50. Gallagher, K. M., Jara, M., Demaria, A. et al., (2003). *The reliability of passively collected AIDS surveillance data in Massachusetts*. Annals of Epidemiology 13: 100-104.

51. Salerno, D. F., Franzblau, A., Armstrong, T. J. et al., (2001). *Test-retest reliability of the upper extremity questionnaire among keyboard operators*. American Journal of Industrial Medicine

40: 655-666.

52. Yokoo, E. M., Valente, J. G., Sichieri, R. et al., (2001). *Validation and calibration of mercury intake through self- referred fish consumption in riverine populations in Pantanal Mato-grossense, Brazil.* Environmental Research 86: 88-93.

53. Xu, L. Z., Porteous, J. E., Phillips, M. R. et al., (2000). *Development and validation of a calcium intake questionnaire for postmenopausal women in China.* Annals of Epidemiology 10: 169-175.

54. Osganian, S. K., Stampfer, M. J., Spiegelman, D. et al., (1999). *Distribution of and factors associated with serum homocysteine levels in children.* Journal of the American Medical Association 281: 1189-1196.

55. Gustafson, P., (2004). *Measurement Error and Misclassification in Statistics and Epidemiology.* Chapman & Hall/CRC, London.

56. Maldonado, G. and Greenland, S., (1997). *The importance of critically interpreting simulation studies.* Epidemiology 8: 453-456.

57. White, I., Frost, C., and Tokunaga, S., (2001). *Correcting for measurement error in binary and continuous variables using replicates.* Statistics in Medicine 20: 3441-3457.

58. Sonis, J., (1998). *A closer look at confounding.* Family Medicine 30: 584-588.

59. Hernán, M. A., Hernández-Díaz, S., Werler, M. M. et al., (2002). *Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology.* American Journal of Epidemiology 155: 176-184.

60. Greenland, S., (1996). *Basic methods for sensitivity analysis of biases.* International Journal of Epidemiology 25: 1107-1116.

61. Lash, T. L. and Fink, A. K., (2003). *Semi-automated sensitivity analysis to assess systematic errors in observational data.* Epidemiology 14: 451-458.

62. Rubin, D. B., (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York.

63. Schafer, J. L., (1997). *Analysis of Incomplete Multivariate Data.* Chapman & Hall, London.

64. Thomas, D., Stram, D., and Dwyer, J., (1993). *Exposure measurement error - influence on exposure-disease relationships and methods of correction.* Annual Review of Public Health 14: 69-93.

65. Carroll, R. J., Ruppert, D., and Stefanski, L. A., (1995). *Measurement Error in Nonlinear Models.* Chapman & Hall/CRC, London.

66. Thurigen, D., Spiegelman, D., Blettner, M. et al., (2000). *Measurement error correction using validation data: a review of methods and their applicability in case-control studies.* Statistical Methods in Medical Research 9: 447-474.

67. Prentice, R. L., (1982). *Covariate measurement errors and parameter-estimation in a failure time regression-model.* Biometrika 69: 331-342.

68. Rosner, B., Willett, W. C., and Spiegelman, D., (1989). *Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error.* Statistics in Medicine 8: 1051-1069.

69. Rosner, B., Spiegelman, D., and Willett, W. C., (1990). *Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error.* American Journal of Epidemiology 132: 734-745.

70. Carroll, R. J. and Stefanski, L. A., (1990). *Approximate quasi-likelihood estimation in models with surrogate predictors.* Journal of the American Statistical Association 85: 652-663.

71. Thurston, S. W., Spiegelman, D., and Ruppert, D., (2003). *Equivalence of regression calibration methods in main study/external validation study designs.* Journal of Statistical Planning and Inference 113: 527-539.

72. Kipnis, V., Carroll, R. J., Freedman, L. S. et al., (1999). *Implications of a new dietary measurement error model for estimation of relative risk: Application to four calibration studies.* American Journal of Epidemiology 150: 642-651.

73. Freedman, L. S., Carroll, R. J., and Wax, Y., (1991). *Estimating the relation between dietary intake obtained from a food frequency questionnaire and true average intake.* American Journal of Epidemiology 134: 310-320.

74. Spiegelman, D., Schneeweiss, S., and McDermott, A., (1997). *Measurement error correction for logistic regression models with an ''alloyed gold standard''.* American Journal of Epidemiology 145: 184-196.

75. Spiegelman, D., Zhao, B., and Kim, J., (2005). *Correlated errors in biased surrogates: study designs and methods for measurement error correction.* Statistics in Medicine 24: 1657-1682.

76. Spiegelman, D., Carroll, R. J., and Kipnis, V., (2001). *Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument.* Statistics in Medicine 20: 139-160.

77. Tosteson, T. D., Stefanski, L. A., and Schafer, D. W., (1989). *A measurement-error model for binary and ordinal regression.* Statistics in Medicine 8: 1139-1147.

78. Gleser, L. J., (1992). *The importance of assessing measurement reliability in multivariate regression.* Journal of the American Statistical Association 87: 696-707.

79. Xie, S. X., Wang, C. Y., and Prentice, R. L., (2001). *A risk set calibration method for failure time regression by using a covariate reliability sample.* Journal of the Royal Statistical Society Series B 63: 855-870.

80. Gorfine, M., Hsu, L., and Prentice, R. L., (2004). *Nonparametric correction for covariate measurement error in a stratified Cox model.* Biostatistics 5: 75-87.

81. Cook, J. R. and Stefanski, L. A., (1994). *Simulation-extrapolation estimation in parametric measurement error models.* Journal of the American Statistical Association 89: 1314-1328.

82. Carroll, R. J., Küchenhoff, H., Lombard, F. et al., (1996). *Asymptotics for the SIMEX estimator in nonlinear measurement error models.* Journal of the American Statistical Association 91: 242-250.

83. Fung, K. Y. and Krewski, D., (1999). *Evaluation of regression calibration and SIMEX methods in logistic regression when one of the predictors is subject to measurement error.* Journal of Epidemiology and Biostatisics 4: 65-74.

84. Robins, J. M., Rotnitzky, A., and Zhao, L. P., (1994). *Estimation of regression coefficients when some regressors are not always observed.* Journal of the American Statistical Association 89: 846-866.

85. Robins, J. M., Hsieh, F. S., and Newey, W., (1995). *Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates.* Journal of the Royal Statistical Society Series B 57: 409-424.

86. Spiegelman, D. and Casella, M., (1997). *Fully parametric and semi-parametric regression*

models for common events with covariate measurement error in main study validation study designs. Biometrics 53: 395–409.

87. Sturmer, T., Thurigen, D., Spiegelman, D. et al., (2002). *The performance of methods for correcting measurement error in case-control studies.* Epidemiology 13: 507-516.

88. Roeder, K., Carroll, R. J., and Lindsay, B. G., (1996). *A semiparametric mixture approach to case-control studies with errors in covariables.* Journal of the American Statistical Association 91: 722-732.

89. Nakamura, T., (1990). *Corrected score function for errors-in-variables models: Methodology and application to generalized linear models.* Biometrika 77: 127-137.

90. Nakamura, T., (1992). *Proportional hazards model with covariates subject to measurement error.* Biometrics 48: 829-838.

91. Huang, Y. J. and Wang, C. Y., (2000). *Cox regression with accurate covariates unascertainable: A nonparametric-correction approach.* Journal of the American Statistical Association 95: 1209-1219.

92. Hu, C. C. and Lin, D. Y., (2002). *Cox regression with covariate measurement error.* Scandinavian Journal of Statistics 29: 637-655.

93. McShane, L. M., Midthune, D. N., Dorgan, J. F. et al., (2001). *Covariate measurement error adjustment for matched case-control studies.* Biometrics 57: 62-73.

94. Carroll, R. J., Gail, M. H., and Lubin, J. H., (1993). *Case-control studies with errors in covariates.* Journal of the American Statistical Association 88: 185-199.

95. Forbes, A. B. and Santner, T. J., (1995). *Estimators of odds ratio regression parameters in matched case-control studies with covariate measurement error.* Journal of the American Statistical Association 90: 1075-1084.

96. Küchenhoff, H. and Carroll, R. J., (1997). *Segmented regression with errors in predictors: Semi-parametric and parametric methods.* Statistics in Medicine 16: 169-188.

97. Hu, P., Tsiatis, A. A., and Davidian, M., (1998). *Estimating the parameters in the Cox model when covariate variables are measured with error.* Biometrics 54: 1407-1419.

98. Liu, K., Mazumdar, S., Stone, R. A. et al., (2001). *Accounting for covariate measurement error in a Cox model analysis of recurrence of depression.* Journal of Psychiatric Research 35: 177-185.

99. Kosinski, A. S. and Flanders, W. D., (1999). *Evaluating the exposure and disease relationship with adjustment for different types of exposure misclassification: A regression approach.* Statistics in Medicine 18: 2795-2808.

100. Spiegelman, D., Rosner, B., and Logan, R., (2000). *Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs.* Journal of the American Statistical Association 95: 51-61.

101. Roy, S., Banerjee, T., and Tapabrata, M., (2005). *Measurement error model for misclassified binary responses.* Statistics in Medicine 24: 269-283.

102. Muller, P. and Roeder, K., (1997). *A Bayesian semiparametric model for case-control studies with errors in variables.* Biometrika 84: 523-537.

103. Gustafson, P., Le, N. D., and Vallee, M., (2000). *Parametric Bayesian analysis of case-control data with imprecise exposure measurements.* Statistics & Probability Letters 47: 357-363.

104. Gustafson, P., Le, N. D., and Vallee, M., (2002). *A Bayesian approach to case-control studies with errors in covariables.* Biostatistics 3: 229-243.

105. Rice, K., (2003). *Full-likelihood approaches to misclassification of a binary exposure in matched case-control studies.* Statistics in Medicine 22: 3177-3194.

106. Prescott, G. J. and Garthwaite, P. H., (2005). *Bayesian analysis of misclassified binary data from a matched case-control study with a validation sub-study.* Statistics in Medicine 24: 379-401.

107. Richardson, S., Leblond, L., Jaussent, I. et al., (2002). *Mixture models in measurement error problems, with reference to epidemiological studies.* Journal of the Royal Statistical Society Series A 165: 549-566.

108. Barron, B. A., (1977). *The effects of misclassification on the estimation of relative risk.* Biometrics 33: 414-418.

109. Copeland, K. T., Checkoway, H., McMichael, A. J. et al., (1977). *Bias due to misclassification in the estimation of relative risk.* American Journal of Epidemiology 105: 488-495.

110. Greenland, S., (1982). *The effect of misclassification in matched-pair case-control studies.* American Journal of Epidemiology 116: 402-406.

111. Greenland, S. and Kleinbaum, D. G., (1983). *Correcting for misclassification in two-way tables and matched-pair studies.* International Journal of Epidemiology 12: 93-97.

112. Greenland, S., (1989). *On correcting for misclassification in twin studies and other matched-pair studies.* Statistics in Medicine 8: 825-829.

113. Duffy, S. W., Rohan, T. E., Kandel, R. et al., (2003). *Misclassification in a matched case-control study with variable matching ratio: application to a study of c-erbB-2 overexpression and breast cancer.* Statistics in Medicine 22: 2459-2468.

114. Espeland, M. A. and Hui, S. L., (1987). *A general approach to analyzing epidemiologic data that contain misclassification errors.* Biometrics 43: 1001-1012.

115. Weinkam, J. J., Rosenbaum, W. L., and Sterling, T. D., (1999). *Recovering true risks when multilevel exposure and covariables are both misclassified.* American Journal of Epidemiology 150: 886-891.

116. Savitz, D. A. and Barón, A. E., (1989). *Estimating and correcting for confounder misclassification.* American Journal of Epidemiology 129: 1062-1071.

117. Liu, X. and Liang, K., (1991). *Adjustment for non-differential misclassification error in the generalized linear model.* Statistics in Medicine 10: 1197-1211.

118. Emsley, C. L., Gao, S., Hall, K. S. et al., (2000). *Estimating odds ratios adjusting for misclassification in Alzheimer's disease risk factor assessment.* Statistics in Medicine 19: 1523-1530.

119. Reade-Christopher, S. J. and Kupper, L. L., (1991). *Effects of exposure misclassification on regression analyses of epidemiologic follow-up study data.* Biometrics 47: 535-548.

120. Wang, C. Y. and Pepe, M. S., (2000). *Expected estimating equations to accommodate covariate measurement error.* Journal of the Royal Statistical Society Series B 62: 509-524.

121. Cornfield, J., Haenszel, W., Cuyler Hammond, E. et al., (1959). *Smoking and lung cancer: Recent evidence and a discussion of some questions.* Journal of the National Cancer Institute 22: 173-203.

122. Bross, I. D. J., (1966). *Spurious effects from an extraneous variable.* Journal of Chronic Diseases 19: 637-647.

123. Bross, I. D. J., (1967). *Pertinency of an extraneous variable.* Journal of Chronic Diseases 20: 487-495.

124. Schlesselman, J. J., (1978). *Assessing effects of confounding variables.* American Journal of Epidemiology 108: 3-8.

125. Simon, R., (1980). *Re: "Assessing effects of confounding variables" (Letter).* American Journal of Epidemiology 111: 127-128.

126. Rosenbaum, P. R. and Rubin, D. B., (1983). *Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome.* Journal of the Royal Statistical Society Series B 45: 212-218.

127. Yanagawa, T., (1984). *Case-control studies: Assessing the effect of a confounding factor.* Biometrika 71: 191-194.

128. Lin, D. Y., Psaty, B. M., and Kronmal, R. A., (1998). *Assessing the sensitivity of regression results to unmeasured confounders in observational studies.* Biometrics 54: 948-963.

129. Stürmer, T., Schneeweiss, S., Avorn, J. et al., (2005). *Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration.* American Journal of Epidemiology 162: 279-289.

130. Greenland, S., (1987). *Quantitative methods in the review of epidemiologic literature.* Epidemiologic Reviews 9: 1-30.

131. Greenland, S., (1998). Basic methods for sensitivity analysis and external adjustment. In: *Modern Epidemiology,* 2nd edition, Chapter 19, K. J. Rothman and S. Greenland (editors), Lippincott Williams and Wilkins, Philadelphia.

132. Gail, M. H., Wacholder, S., and Lubin, J. H., (1988). *Indirect corrections for confounding under multiplicative and additive risk models.* American Journal of Industrial Medicine 13: 119-130.

133. Greenland, S., (2001). *Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment.* Risk Analysis 21: 579-583.

134. Fox, M. P., Lash, T. L., and Greenland, S., (2005). *A method to automate probabilistic sensitivity analyses of misclassified binary variables.* International Journal of Epidemiology 34: 1370-1376.

135. Lash, T. L. and Silliman, R. A., (2000). *A sensitivity analysis to separate bias due to confounding from bias due to predicting misclassification by a variable that does both.* Epidemiology 11: 544-549.

136. Phillips, C. V., (2003). *Quantifying and reporting uncertainty from systematic errors.* Epidemiology 14: 459-466.

137. Rosenbaum, P. R., (1987). *Sensitivity analysis for certain permutation inferences in matched observational studies.* Biometrika 74: 13-26.

138. Rosenbaum, P. R., (1988). *Sensitivity analysis for matching with multiple controls.* Biometrika 75: 577-581.

139. Rosenbaum, P. R., (1991). *Sensitivity analysis for matched case-control studies.* Biometrics 47: 87-100.

140. Rosenbaum, P. R., (1997). *Signed rank statistics for coherent predictions.* Biometrics 53: 556-566.

141. Rosenbaum, P. R. and Krieger, A. M., (1990). *Sensitivity of two-sample permutation inferences in observational studies.* Journal of the American Statistical Association 85: 493-498.

142. Jo, B., (2002). *Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance.* Statistics in Medicine 21: 3161-3181.

143. Angrist, J. D., Imbens, G. W., and Rubin, D. B., (1996). *Identification of causal effects using instrumental variables.* Journal of the American Statistical Association 91: 444-455.

144. Hardin, J. W. and Schmiediche, H., (2003). *The simulation extrapolation method for fitting generalized linear models with additive measurement error.* The Stata Journal 3: 373-385.

145. Hardin, J. W., Schmiediche, H., and Carroll, R. J., (2003). *The regression-calibration method for fitting generalized linear models with additive measurement error.* The Stata Journal 3: 361-372.

146. Davey Smith, G. and Phillips, A. N., (1996). *Inflation in epidemiology: "The proof and measurement of association between two things" revisited.* British Medical Journal 312: 1659-1661.

147. Hirschfield, G. M. and Pepys, M. B., (2003). *C-reactive protein and cardiovascular disease: new insights from an old molecule.* QJM 96: 793-807.

148. Rifai, N. and Ridker, P. M., (2001). *High-sensitivity C-reactive protein: A novel and promising marker of coronary heart disease.* Clinical Chemistry 47: 403-411.

149. Onat, A., Sansoy, V., Yildirim, B. et al., (2001). *C-reactive protein and coronary heart disease in western Turkey.* American Journal of Cardiology 88: 601-607.

150. Jousilahti, P., Salomaa, V., Rasi, V. et al., (2001). *The association of C-reactive protein, serum amyloid a and fibrinogen with prevalent coronary heart disease - baseline findings of the PAIS project.* Atherosclerosis 156: 451-456.

151. Woodward, M., Rumley, A., Lowe, G. D. O. et al., (2003). *C-reactive protein: associations with haematological variables, cardiovascular risk factors and prevalent cardiovascular disease.* British Journal of Haematology 122: 135-141.

152. Anand, S. S., Razak, F., Yi, Q. et al., (2004). *C-reactive protein as a screening test for cardiovascular risk in a multiethnic population.* Arteriosclerosis, Thrombosis, and Vascular Biology 24: 1509-1515.

153. Lawlor, D. A., Davey Smith, G., Rumley, A. et al., (2005). *Associations of fibrinogen and C-reactive protein with prevalent and incident coronary heart disease are attenuated by adjustment for confounding factors: British Women's Heart and Health study.* Thrombosis and Haemostasis 93: 955-963.

154. Kuller, L. H., Tracy, R. P., Shaten, J. et al., (1996). *Relation of C-reactive protein and coronary heart disease in the MRFIT nested case-control study.* American Journal of Epidemiology 144: 537-547.

155. Tracy, R. P., Lemaitre, R. N., Psaty, B. M. et al., (1997). *Relationship of C-reactive protein to risk of cardiovascular disease in the elderly: Results from the Cardiovascular Health Study and the Rural Health Promotion Project.* Arteriosclerosis Thrombosis and Vascular Biology 17: 1121-1127.

156. Ridker, P. M., Cushman, M., Stampfer, M. J. et al., (1997). *Inflammation, aspirin, and the risk*

of cardiovascular disease in apparently healthy men. New England Journal of Medicine 336: 973-979.

157. Ridker, P. M., Buring, J. E., Shih, J. et al., (1998). *Prospective study of C-reactive protein and the risk of future cardiovascular events among apparently healthy women.* Circulation 98: 731-733.

158. Ridker, P. M., Hennekens, C. H., Buring, J. E. et al., (2000). *C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women.* New England Journal of Medicine 342: 836-843.

159. Packard, C. J., O'Reilly, D. S., Caslake, M. J. et al., (2000). *Lipoprotein-associated phospholipase A_2 as an independent predictor of coronary heart disease.* New England Journal of Medicine 343: 1148-1155.

160. Danesh, J., Whincup, P., Walker, M. et al., (2000). *Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses.* British Medical Journal 321: 199-204.

161. Roivainen, M., Viik-Kajander, M., Palosuo, T. et al., (2000). *Infections, inflammation, and the risk of coronary heart disease.* Circulation 101: 252-257.

162. Pradhan, A. D., Manson, J. E., Rossouw, J. E. et al., (2002). *Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: Prospective analysis from the Women's Health Initiative Observational Study.* Journal of the American Medical Association 288: 980-987.

163. van der Meer, I. M., de Maat, M. P. M., Kiliaan, A. J. et al., (2003). *The value of C-reactive protein in cardiovascular risk prediction: The Rotterdam Study.* Archives of Internal Medicine 163: 1323-1328.

164. Luc, G., Bard, J. M., Juhan-Vague, I. et al., (2003). *C-reactive protein, interleukin-6, and fibrinogen as predictors of coronary heart disease. The PRIME study.* Arteriosclerosis Thrombosis and Vascular Biology 23: 1255-1261.

165. Danesh, J., Wheeler, J. G., Hirschfield, G. M. et al., (2004). *C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease.* New England Journal of Medicine 350: 1387-1397.

166. Pai, J. K., Pischon, T., Ma, J. et al., (2004). *Inflammatory markers and the risk of coronary heart disease in men and women.* New England Journal of Medicine 351: 2599-2610.

167. Folsom, A. R., Aleksic, N., Catellier, D. et al., (2002). *C-Reactive protein and incident coronary heart disease in the Atherosclerosis Risk In Communities (ARIC) study.* American Heart Journal 144: 233-238.

168. Ballantyne, C. M., Hoogeveen, R. C., Bang, H. et al., (2004). *Lipoprotein-associated phospholipase A_2, high-sensitivity C-reactive protein, and risk for incident coronary heart disease in middle-aged men and women in the Atherosclerosis Risk in Communities (ARIC) study.* Circulation 109: 837-842.

169. Koenig, W., Sund, M., Fröhlich, M. et al., (1999). *C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: Results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992.* Circulation 99: 237-242.

170. Strandberg, T. E. and Tilvis, R. S., (2000). *C-reactive protein, cardiovascular risk factors, and mortality in a prospective study in the elderly.* Arteriosclerosis, Thrombosis, and Vascular Biology 20: 1057-1060.

171. Mendall, M. A., Strachan, D. P., Butland, B. K. et al., (2000). *C-reactive protein: relation to total mortality, cardiovascular mortality and cardiovascular risk factors in men.* European Heart Journal 21: 1584-1590.

172. Lowe, G. D. O., Yarnell, J. W. G., Rumley, A. et al., (2001). *C-reactive protein, fibrin D-dimer, and incident ischemic heart disease in the Speedwell Study: Are inflammation and fibrin turnover linked in pathogenesis?* Arteriosclerosis Thrombosis and Vascular Biology 21: 603-610.

173. Ridker, P. M., Rifai, N., Rose, L. et al., (2002). *Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events.* New England Journal of Medicine 347: 1557-1565.

174. Blake, G. J., Rifai, N., Buring, J. E. et al., (2003). *Blood pressure, C-reactive protein, and risk of future cardiovascular events.* Circulation 108: 2993-2999.

175. Koenig, W., Sund, M., Fröhlich, M. et al., (2003). *Refinement of the association of serum C-reactive protein concentration and coronary heart disease risk by correction for within-subject variation over time.* American Journal of Epidemiology 158: 357-364.

176. Cesari, M., Penninx, B. W. J. H., Newman, A. B. et al., (2003). *Inflammatory markers and cardiovascular disease (The Health, Aging and Body Composition (Health ABC) Study).* American Journal of Cardiology 92: 522-528.

177. Cesari, M., Penninx, B. W. J. H., Newman, A. B. et al., (2003). *Inflammatory markers and the onset of cardiovascular events: Results from the Health ABC Study.* Circulation 108: 2317-2322.

178. Lowe, G. D. O., Sweetnam, P. M., Yarnell, J. W. G. et al., (2004). *C-reactive protein, fibrin D-dimer, and risk of ischemic heart disease: The Caerphilly and Speedwell Studies.* Arteriosclerosis, Thrombosis, and Vascular Biology 24: 1957-1962.

179. Best, L. G., Zhang, Y., Lee, E. T. et al., (2005). *C-reactive protein as a predictor of cardiovascular risk in a population with a high prevalence of diabetes.* Circulation 112: 1289-1295.

180. St-Pierre, A. C., Cantin, B., Bergeron, J. et al., (2005). *Inflammatory markers and long-term risk of ischemic heart disease in men. A 13-year follow-up of the Québec Cardiovascular Study.* Atherosclerosis 182: 315-321.

181. Ridker, P. M., Rifai, N., Cook, N. R. et al., (2005). *Non-HDL cholesterol, apolipoproteins A-1 and B_{100}, standard lipid measures, lipid ratios, and CRP as risk factors for cardiovascular disease in women.* Journal of the American Medical Association 294: 326-333.

182. Laaksonen, D. E., Niskanen, L., Nyyssönen, K. et al., (2005). *C-reactive protein in the prediction of cardiovascular and overall mortality in middle-aged men: a population-based cohort study.* European Heart Journal 26: 1783-1789.

183. Cushman, M., Arnold, A. M., Psaty, B. M. et al., (2005). *C-reactive protein and the 10-year incidence of coronary heart disease in older men and women. The Cardiovascular Health Study.* Circulation 112: 25-31.

184. Danesh, J., Collins, R., Appleby, P. et al., (1998). *Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: Meta-analyses of prospective studies.* Journal of the American Medical Association 279: 1477-1482.

185. Davey Smith, G., Harbord, R., and Ebrahim, S., (2004). *Fibrinogen, C-reactive protein and coronary heart disease: does Mendelian randomization suggest the associations are non- causal?* Quarterly Journal of Medicine 97: 163-166.

186. Davey Smith, G. and Ebrahim, S., (2003). *'Mendelian randomization': Can genetic*

epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology 32: 1-22.

187. Zee, R. Y. L. and Ridker, P. M., (2002). *Polymorphism in the human C-reactive protein (CRP) gene, plasma concentrations of CRP, and the risk of future arterial thrombosis.* Atherosclerosis 162: 217-219.

188. Casas, J. P., Shah, T., Cooper, J. et al., (in press). *Insight into the nature of the CRP-coronary event association using Mendelian randomization.* International Journal of Epidemiology.

189. Lawlor, D. A., Ebrahim, S., and Davey Smith, G., (2002). *Socioeconomic position in childhood and adulthood and insulin resistance: cross sectional survey using data from the British women's heart and health study.* British Medical Journal 325: 805-807.

190. Lawlor, D. A., Bedford, C., Taylor, M. et al., (2003). *Geographical variation in cardiovascular disease, risk factors, and their control in older women: British Women's Heart and Health Study.* Journal of Epidemiology and Community Health 57: 134-140.

191. Lebowitz, M., Knudson, R., Roberson, G. et al., (1982). *Significance of intraindividual changes in maximum expiratory flow volume and peak expiratory flow measurement.* Chest 81: 566-570.

192. Rozas, C. and Goldman, A., (1982). *Daily spirometric variability; normal subjects and subjects with chronic bronchitis with and without airflow obstruction.* Archives of Internal Medicine 142: 1287-1291.

193. Groth, S., Dirksen, A., Dirksen, H. et al., (1986). *Intraindividual variation and effect of learning in lung function examinations: a population study.* Bulletin Europeen de Physiopathologie Respiratoire 22: 35-42.

194. Randell, J. T., Salonen, R. O., Pekkarinen, H. et al., (1999). *Short-term variations in oscillatory and spirometric lung function parameters of non-asthmatic adults.* Clinical Physiology 19: 329-337.

195. Jacobs, D. R. and Barrett-Connor, E., (1982). *Retest reliability of plasma cholesterol and triglyceride - the Lipid Research Clinics Prevalence Study.* American Journal of Epidemiology 116: 878-885.

196. Godsland, I., (1985). *Intra-individual variation: significant changes in parameters of lipid and carbohydrate metabolism in the individual and intra-individual variation in different test populations.* Annals of Clinical Biochemistry 22: 618-624.

197. Bookstein, L., Gidding, S. S., Donovan, M. et al., (1990). *Day-to-day variability of serum cholesterol, triglyceride, and high-density lipoprotein cholesterol levels. Impact on the assessment of risk according to the National Cholesterol Education Program guidelines.* Archives of Internal Medicine 150: 1653-1657.

198. Brenner, H. and Heiss, G., (1990). *The intraindividual variability of fasting triglyceride - a challenge for further standardization.* European Heart Journal 11: 1054-1058.

199. Marcovina, S. M., Gaur, V. P., and Albers, J. J., (1994). *Biological variability of cholesterol, triglyceride, low- and high-density lipoprotein cholesterol, lipoprotein(a), and apolipoproteins A-I and B.* Clinical Chemistry 40: 574-578.

200. Egger, M., Davey Smith, G., Pfluger, D. et al., (1999). *Triglyceride as a risk factor for ischaemic heart disease in British men: effect of adjusting for measurement error.* Atherosclerosis 143: 275-284.

201. Clark, G. H. and Fraser, C. G., (1993). *Biological variation of acute phase proteins.* Annals of Clinical Biochemistry 30: 373-376.

202. de Maat, M. P. M., de Bart, A. C. W., Hennis, B. C. et al., (1996). *Interindividual and intraindividual variability in plasma fibrinogen, TPA antigen, PAI activity, and CRP in healthy, young volunteers and patients with angina pectoris.* Arteriosclerosis, Thrombosis, and Vascular Biology 16: 1156-1162.

203. de Maat, M. P. M., Arnold, A. E. R., van Buuren, S. et al., (1996). *Modulation of plasma fibrinogen levels by ticlopidine in healthy volunteers and patients with stable angina pectoris.* Thrombosis and Haemostasis 76: 166-170.

204. Schuit, A. J., Schouten, E. G., Kluft, C. et al., (1997). *Effect of strenuous exercise on fibrinogen and fibrinolysis in healthy elderly men and women.* Thrombosis and Haemostasis 78: 845-851.

205. Macy, E. M., Hayes, T. E., and Tracy, R. P., (1997). *Variability in the measurement of C-reactive protein in healthy subjects: implications for reference intervals and epidemiological applications.* Clinical Chemistry 43: 52-58.

206. Sakkinen, P. A., Macy, E. M., Callas, P. W. et al., (1999). *Analytical and biological variability in measures of hemostasis, fibrinolysis, and inflammation: Assessment and implications for epidemiology.* American Journal of Epidemiology 149: 261-267.

207. Riese, H., (2000). *Job strain and risk for cardiovascular disease in female nurses.* Thesis, Universiteit van Amsterdam, The Netherlands.

208. Ockene, I. S., Matthews, C. E., Rifai, N. et al., (2001). *Variability and classification accuracy of serial high-sensitivity C-reactive protein measurements in healthy adults.* Clinical Chemistry 43: 444-450.

209. Koenig, W., Sund, M., Fröhlich, M. et al., (2003). *Refinement of the association of serum C-reactive protein concentration and coronary heart disease risk by correction for within-subject variation over time: The MONICA Augsburg Studies, 1984-1987.* American Journal of Epidemiology 158: 357-364.

210. Kluft, C. and de Maat, M. P. M., (2001). *Determination of the habitual low blood level of C-reactive protein in individuals.* Italian Heart Journal 2: 172-180.

211. White, I. R., (2005). *Uses and limitations of randomization-based efficacy estimators.* Statistical Methods in Medical Research 14: 327-347.

212. Greenland, S., (2000). *An introduction to instrumental variables for epidemiologists.* International Journal of Epidemiology 29: 722-729.

213. Holland, P. W., (1986). *Statistics and causal inference.* Journal of the American Statistical Association 81: 945-960.

214. Rubin, D. B., (1974). *Estimating causal effects of treatments in randomized and nonrandomized studies.* Journal of Educational Psychology 66: 688-701.

215. Rubin, D. B., (1977). *Assignment of treatment group on the basis of a covariate.* Journal of Educational Statistics 2: 1-26.

216. Rubin, D. B., (1978). *Bayesian inference for causal effects: the role of randomization.* Annals of Statistics 7: 34-58.

217. Rubin, D. B., (1980). *Discussion of "Randomization analysis of experimental data: the Fisher randomization test" by D. Basu.* Journal of the American Statistical Association 75: 591-593.

218. Frangakis, C. E., Rubin, D. B., and Zhou, X. H., (2002). *Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms.* Biostatistics 3: 147-164.

219. Sommer, A. and Zeger, S. L., (1991). *On estimating efficacy from clinical trials.* Statistics in Medicine 10: 45-52.

220. Albert, J. M., (2002). *Estimating efficacy in clinical trials with clustered binary responses.* Statistics in Medicine 21: 649-661.

221. Cuzick, J., Edwards, R., and Segnan, N., (1997). *Adjusting for non-compliance and contamination in randomized clinical trials.* Statistics in Medicine 16: 1017-1029.

222. Branson, M. and Whitehead, J., (2003). *A score test for binary data with patient non-compliance.* Statistics in Medicine 22: 3115-3132.

223. Baker, S. G., (1998). *Analysis of survival data from a randomized trial with all-or-none compliance: Estimating the cost-effectiveness of a cancer screening program.* Journal of the American Statistical Association 93: 929-934.

224. Imbens, G. W. and Rubin, D. B., (1997). *Bayesian inference for causal effects in randomized experiments with noncompliance.* Annals of Statistics 25: 305-327.

225. Hirano, K., Imbens, G. W., Rubin, D. et al., (2000). *Assessing the effect of an influenza vaccine in an encouragement design.* Biostatistics 1: 69-88.

226. Frangakis, C. E. and Rubin, D. B., (1999). *Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes.* Biometrika 86: 365-379.

227. Levy, D. E., O'Malley, A. J., and Normand, S. L. T., (2004). *Covariate adjustment in clinical trials with non-ignorable missing data and non-compliance.* Statistics in Medicine 23: 2319-2339.

228. Loeys, T. and Goetghebeur, E., (2003). *A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance.* Biometrics 59: 100-105.

229. Kim, L. G. and White, I. R., (2004). *Compliance-adjusted intervention effects in survival data.* The Stata Journal 4: 257-264.

230. Ten Have, T. R., Joffe, M., and Cary, M., (2003). *Causal logistic models for non-compliance under randomized treatment with univariate binary response.* Statistics in Medicine 22: 1255-1283.

231. Nagelkerke, N., Fidler, V., Bernsen, R. et al., (2000). *Estimating treatment effects in randomized clinical trials in the presence of non-compliance.* Statistics in Medicine 19: 1849-1864.

232. Efron, B. and Feldman, D., (1991). *Compliance as an explanatory variable in clinical trials.* Journal of the American Statistical Association 86: 9-17.

233. Zeger, S. L. and Liang, K., (1991). *Dose-response estimands.* Journal of the American Statistical Association 86: 18-19.

234. Dunn, G., (1999). *The problem of measurement error in modelling the effect of compliance in a randomized trial.* Statistics in Medicine 18: 2863-2877.

235. Goetghebeur, E. and Molenberghs, G., (1996). *Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance.* Journal of the American Statistical Association 91: 928-934.

236. Albert, J. M., (1999). *A threshold causal model for clinical trials with departures from intended*

treatment. Statistics in Medicine 18: 1615-1626.

237. Robins, J. M. and Tsiatis, A. A., (1991). *Correcting for non-compliance in randomized trials using rank preserving structural failure time models*. Communications in Statistics Theory and Methods 20: 2609-2631.

238. White, I. R., Babiker, A. G., Walker, S. et al., (1999). *Randomization-based methods for correcting for treatment changes: Examples from the Concorde Trial*. Statistics in Medicine 18: 2617-2634.

239. Hernán, M. A., Cole, S. R., Margolick, J. et al., (2005). *Structural accelerated failure time models for survival analysis in studies with time-varying treatments*. Pharmacoepidemiology and Drug Safety 14: 477-491.

240. Walker, A. S., White, I. R., and Babiker, A. G., (2004). *Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment*. Statistics in Medicine 23: 571-590.

241. White, I. R. and Walker, S., (2002). *strbee: Randomization-based efficacy estimator*. The Stata Journal 2: 140-150.

242. Mark, S. D. and Robins, J. M., (1993). *Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model*. Statistics in Medicine 12: 1605-1628.

243. White, I. R., Walker, S., Babiker, A. G. et al., (1997). *Impact of treatment changes on the interpretation of the Concorde trial*. AIDS 11: 999-1006.

244. Mark, S. D. and Robins, J. M., (1993). *A method for the analysis of randomized trials with compliance information: An application to the Multiple Risk Factor Intervention Trial*. Controlled Clinical Trials 14: 79-97.

245. Korhonen, P. A., Laird, N. M., and Palmgren, J., (1999). *Correcting for non-compliance in randomized trials: An application to the ATBC Study*. Statistics in Medicine 18: 2879-2897.

246. Korhonen, P., Loeys, T., Goetghebeur, E. et al., (2000). *Vitamin A and infant mortality: beyond intention-to-treat in a randomized trial*. Lifetime Data Analysis 6: 107-121.

247. Loeys, T., Vansteelandt, S., and Goetghebeur, E., (2001). *Accounting for correlation and compliance in cluster randomized trials*. Statistics in Medicine 20: 3753-3767.

248. Loeys, T. and Goetghebeur, E., (2002). *Baseline information in structural failure time estimators for the effect of observed treatment compliance*. Statistics in Medicine 21: 1173-1188.

249. Matsui, S., (2004). *Analysis of times to repeated events in two-arm randomized trials with noncompliance and dependent censoring*. Biometrics 60: 965-976.

250. Vandebosch, A., Goetghebeur, E., and Van Damme, L., (2005). *Structural accelerated failure time models for the effects of observed exposures on repeated events in a clinical trial*. Statistics in Medicine 24: 1029-1046.

251. Wei, L. J., Lin, D. Y., and Weissfeld, L., (1989). *Regression analysis of multivariate incomplete failure time data by modeling marginal distributions*. Journal of the American Statistical Association 84: 1065-1073.

252. Sato, T., (2001). *A method for the analysis of repeated binary outcomes in randomized clinical trials with non-compliance*. Statistics in Medicine 20: 2761-2774.

253. Robins, J. M. and Finkelstein, D. M., (2000). *Correcting for noncompliance and dependent*

censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 56: 779-788.

254. Robins, J. M., (1994). *Correcting for noncompliance in randomized trials using structural nested mean models.* Communications in Statistics-Theory and Methods 23: 2379-2412.

255. Robins, J. M., (1989). The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: *Health Service Research Methodology: A Focus on AIDS,* L. Sechrest, H. Freeman, and A. Mulley (editors), U.S. Public Health Service, National Center for Health Services Research, Washington D.C.

256. Goetghebeur, E. and Vansteelandt, S., (2005). *Structural mean models for compliance analysis in randomized clinical trials and the impact of errors on measures of exposure.* Statistical Methods in Medical Research 14: 397-415.

257. Goetghebeur, E. and Lapp, K., (1997). *The effect of treatment compliance in a placebo-controlled trial: Regression with unpaired data.* Journal of the Royal Statistical Society Series C 46: 351-364.

258. Fischer-Lapp, K. and Goetghebeur, E., (1999). *Practical properties of some structural mean analyses of the effect of compliance in randomized trials.* Controlled Clinical Trials 20: 531-546.

259. Matsuyama, Y., (2002). *Correcting for non-compliance of repeated binary outcomes in randomized clinical trials: randomized analysis approach.* Statistics in Medicine 21: 675-687.

260. Joffe, M. and Bresinger, C., (2003). *Weighting in instrumental variables and G-estimation.* Statistics in Medicine 22: 1285-1303.

261. Gawkrodger, D. J., (2002). Malignant epidermal tumours. In: *Dermatology: An Illustrated Colour Text,* 3rd edition, Elsevier Health Sciences, Edinburgh.

262. Zak-Prelich, M., Narbutt, J., and Sysa-Jedrzejowska, A., (2004). *Environmental risk factors predisposing to the development of basal cell carcinoma.* Dermatologic Surgery 30: 248-252.

263. English, D. R., Armstrong, B. K., Kricker, A. et al., (1997). *Sunlight and cancer.* Cancer Causes & Control 8: 271-283.

264. Hunter, D. J., Colditz, G. A., Stampfer, M. J. et al., (1990). *Risk factors for basal cell carcinoma in a prospective cohort of women.* Annals of Epidemiology 1: 13-23.

265. Marks, R., Staples, M., and Giles, G. G., (1993). *Trends in non-melanocytic skin cancer treated in Australia: The second national survey.* International Journal of Cancer 53: 585-590.

266. Kricker, A., Armstrong, B. K., English, D. R. et al., (1995). *A dose-response curve for sun exposure and basal cell carcinoma.* International Journal of Cancer 60: 482-488.

267. van Dam, R. M., Huang, Z., Rimm, E. B. et al., (1999). *Risk factors for basal cell carcinoma of the skin in men: Results from the Health Professionals Follow-up Study.* American Journal of Epidemiology 150: 459-468.

268. Marks, R., Jolley, D., Dorevitch, A. P. et al., (1989). *The incidence of non-melanocytic skin cancers in an Australian population: results of a 5-year prospective study.* Medical Journal of Australia 150: 475-478.

269. Vitasa, B. C., Taylor, H. R., Strickland, P. T. et al., (1990). *Association of nonmelanoma skin cancer and actinic keratosis with cumulative solar ultraviolet exposure in Maryland watermen.* Cancer 65: 2811-2817.

270. Kricker, A., Armstrong, B. K., English, D. R. et al., (1991). *Pigmentary and cutaneous risk factors for non-melanocytic skin cancer - a case-control study.* International Journal of Cancer 48: 650-662.

271. Gallagher, R. P., Hill, G. B., Bajdik, C. D. et al., (1995). *Sunlight exposure, pigmentary factors, and risk of nonmelanocytic skin cancer. 1. Basal cell carcinoma.* Archives of Dermatology 131: 157-163.

272. Zanetti, R., Rosso, S., Martinez, C. et al., (1996). *The multicentre south European study 'Helios' .1. Skin characteristics and sunburns in basal cell and squamous cell carcinomas of the skin.* British Journal of Cancer 73: 1440-1446.

273. Naldi, L., DiLandro, A., D'Avanzo, B. et al., (2000). *Host-related and environmental risk factors for cutaneous basal cell carcinoma: Evidence from an Italian case-control study.* Journal of the American Academy of Dermatology 42: 446-452.

274. Vlajinac, H. D., Adanja, B. J., Lazer, Z. F. et al., (2000). *Risk factors for basal cell carcinoma.* Acta Oncologica 39: 611-616.

275. Suárez-Varela, M. M., Llopis-Gonzalez, A., and Ferrer-Caraco, E., (1996). *Non-melanoma skin cancer: a case-control study on risk factors and protective measures.* Journal of Environmental Pathology, Toxicology and Oncology 15: 255-261.

276. Marks, R., Ponsford, M. W., Selwood, T. S. et al., (1983). *Non-melanotic skin cancer and solar keratoses in Victoria.* Medical Journal of Australia 2: 619-622.

277. Walther, U., Kron, M., Sander, S. et al., (2004). *Risk and protective factors for sporadic basal cell carcinoma: results of a two-centre case-control study in southern Germany. Clinical actinic elastosis may be a protective factor.* British Journal of Dermatology 151: 170-178.

278. Green, A., Williams, G., Neale, R. et al., (1999). *Daily sunscreen application and betacarotene supplementation in prevention of basal-cell and squamous-cell carcinomas of the skin: a randomised controlled trial.* Lancet 354: 723-729.

279. Kricker, A., Armstrong, B. K., English, D. R. et al., (1995). *Does intermittent sun exposure cause basal cell carcinoma? A case-control study in Western Australia.* International Journal of Cancer 60: 489-494.

280. Armstrong, B. K. and Kricker, A., (2001). *The epidemiology of UV induced skin cancer.* Journal of Photochemistry and Photobiology.B, Biology 63: 8-18.

281. Rosso, S., Zanetti, R., Martinez, C. et al., (1996). *The multicentre south European study 'Helios' .2. Different sun exposure patterns in the aetiology of basal cell and squamous cell carcinomas of the skin.* British Journal of Cancer 73: 1447-1454.

282. Green, A. and Battistutta, D., (1990). *Incidence and determinants of skin cancer in a high-risk Australian population.* International Journal of Cancer 46: 356-361.

283. Nakazawa, H., English, D., Randell, P. L. et al., (1994). *UV and skin cancer: Specific p53 gene mutation in normal skin as a biologically relevant exposure measure.* Proceedings of the National Academy of Sciences of the United States of America 91: 360-364.

284. Ziegler, A., Leffell, D. J., Kunala, S. et al., (1993). *Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers.* Proceedings of the National Academy of Sciences of the United States of America 90: 4216-4220.

285. Kantor, J. (29/10/04). Actinic keratosis. *MedlinePlus.* Retrieved 10/05/06, from http://www.nlm.nih.gov/medlineplus/ency/article/000827.htm.

286. Anonymous (19/10/05). What are Actinic Keratoses? *ActinicKeratosesNet.* Retrieved 10/05/06, from http://www.skincarephysicians.com/actinickeratosesnet/whatare.html.

287. Kantor, J. (29/10/04). Liver Spots. *MedlinePlus.* Retrieved 10/05/06, from http://www.nlm.nih.gov/medlineplus/ency/article/001141.htm.

288. Kantor, J. (29/10/04). Birthmarks - pigmented. *MedlinePlus.* Retrieved 10/05/06, from http://www.nlm.nih.gov/medlineplus/ency/article/000828.htm.

289. Anonymous (18/01/06). Aging skin. *Penn State Milton S Hershey Medical Center.* Retrieved 10/05/06, from http://www.hmc.psu.edu/healthinfo/a/agingskin.htm.

290. Kantor, J. (22/07/05). Telangiectasia. *MedlinePlus.* Retrieved 16/05/06, from http://www.nlm.nih.gov/medlineplus/ency/article/003284.htm.

291. Vainio, H., Miller, A. B., and Bianchini, F., (2000). *An international evaluation of the cancer-preventive potential of sunscreens - Meeting held at Lyon, France, 10-18, 2000.* International Journal of Cancer 88: 838-842.

292. Pandeya, N., Purdie, D. M., Green, A. et al., (2005). *Repeated occurrence of basal cell carcinoma of the skin and multifailure survival analysis: follow-up data from the Nambour Skin Cancer Prevention Trial.* American Journal of Epidemiology 161: 748-754.

293. Andersen, P. K. and Gill, R. D., (1982). *Cox's regression model for counting processes: a large sample study.* Annals of Statistics 10: 1100-1120.

294. Prentice, R. L., Williams, B. J., and Peterson, A. V., (1981). *On the regression analysis of multivariate failure time data.* Biometrika 68: 373-379.

295. Thompson, S. C., Jolley, D., and Marks, R., (1993). *Reduction of solar keratoses by regular sunscreen use.* New England Journal of Medicine 329: 1147-1151.

296. Naylor, M. F., Boyd, A., Smith, D. W. et al., (1995). *High sun protection sunscreens in the suppression of actinic neoplasia.* Archives of Dermatology 131: 170-175.

297. Darlington, S., Williams, G., Neale, R. et al., (2003). *A randomized controlled trial to assess sunscreen application and beta carotene supplementation in the prevention of solar keratoses.* Archives of Dermatology 139: 451-455.

298. Green, A., Battistutta, D., Hart, V. et al., (1994). *The Nambour Skin Cancer and Actinic Eye Disease Prevention Trial: Design and baseline characteristics of participants.* Controlled Clinical Trials 15: 512-522.

299. Efron, B. and Tibshirani, R. J., (1993). Better bootstrap confidence intervals. In: *An Introduction to the Bootstrap,* Chapter 14, Chapman & Hall, London.

Appendix 1.

Additional results tables from the analyses in Chapter 3

A1.1. Additional tables for two confounders

Table A1.1: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.1, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 Crude OR	0.1 OR adjusted for Z_1	0.1 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 … 0.75	0.1 … 1	0.3 Crude OR	0.3 OR adjusted for Z_1	0.3 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.3 … 0.75	0.3 … 1	0.5 Crude OR	0.5 OR adjusted for Z_1	0.5 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.5 … 0.75	0.5 … 1
0.1	0.5	1.13	1.10	1.07	1.05	1.03	1.28	1.25	1.14	1.09	1.03	1.46	1.42	1.24	1.14	1.02
	0.75		1.08	1.05	1.03	1.02		1.23	1.13	1.07	1.01		1.41	1.23	1.12	1.01
	1		1.06	1.03	1.02	1.00		1.21	1.11	1.06	1.00		1.39	1.21	1.11	1.00
0.3	0.5	1.28	1.17	1.14	1.13	1.11	1.46	1.34	1.23	1.17	1.11	1.66	1.54	1.35	1.24	1.13
	0.75		1.11	1.09	1.07	1.06		1.28	1.17	1.11	1.06		1.48	1.29	1.18	1.06
	1		1.05	1.03	1.01	1.00		1.22	1.11	1.06	1.00		1.41	1.22	1.12	1.00
0.5	0.5	1.46	1.27	1.24	1.23	1.21	1.66	1.47	1.35	1.29	1.22	1.91	1.72	1.52	1.40	1.27
	0.75		1.16	1.14	1.12	1.11		1.37	1.24	1.18	1.12		1.61	1.40	1.27	1.14
	1		1.05	1.02	1.01	1.00		1.25	1.13	1.06	1.00		1.49	1.27	1.14	1.00

* *OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0056.*

Table A1.2: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.2, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.13	1.09	1.06	1.05	1.03	1.28	1.24	1.13	1.08	1.02	1.45	1.42	1.23	1.12	1.00
	0.75		1.07	1.05	1.03	1.02		1.22	1.12	1.07	1.01		1.40	1.22	1.12	1.00
	1		1.06	1.03	1.02	1.00		1.21	1.11	1.06	1.00		1.38	1.21	1.11	1.00
0.3	0.5	1.28	1.16	1.13	1.12	1.11	1.45	1.32	1.21	1.16	1.10	1.65	1.52	1.33	1.22	1.10
	0.75		1.09	1.08	1.07	1.06		1.26	1.16	1.11	1.05		1.45	1.27	1.16	1.05
	1		1.03	1.02	1.01	1.00		1.19	1.10	1.05	1.00		1.38	1.21	1.11	1.00
0.5	0.5	1.45	1.24	1.23	1.22	1.21	1.65	1.44	1.33	1.27	1.21	1.89	1.68	1.48	1.36	1.23
	0.75		1.13	1.12	1.12	1.11		1.32	1.22	1.16	1.11		1.56	1.36	1.24	1.12
	1		1.00	1.00	1.00	1.00		1.19	1.10	1.05	1.00		1.43	1.23	1.12	1.00

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0055.*

Table A1.3: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.3, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.13	1.09	1.06	1.04	1.03	1.27	1.24	1.13	1.07	1.00	1.44	1.41	1.21	1.10	0.98
	0.75		1.07	1.04	1.03	1.01		1.22	1.12	1.06	1.00		1.39	1.21	1.11	0.99
	1		1.05	1.03	1.01	1.00		1.20	1.11	1.06	1.00		1.37	1.21	1.11	1.00
0.3	0.5	1.27	1.15	1.13	1.12	1.11	1.44	1.31	1.20	1.15	1.09	1.64	1.50	1.31	1.20	1.08
	0.75		1.08	1.07	1.06	1.06		1.24	1.15	1.10	1.05		1.43	1.25	1.15	1.04
	1		1.01	1.00	1.00	1.00		1.17	1.09	1.05	1.00		1.35	1.19	1.10	1.00
0.5	0.5	1.44	1.22	1.21	1.21	1.21	1.64	1.41	1.31	1.25	1.19	1.88	1.64	1.44	1.33	1.20
	0.75		1.09	1.10	1.11	1.11		1.28	1.20	1.15	1.10		1.51	1.33	1.22	1.10
	1		0.96	0.98	0.99	1.00		1.14	1.08	1.04	1.00		1.36	1.20	1.10	1.00

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0052.*

Table A1.4: Geometric means of the estimated exposure-outcome odds ratios for the case with two confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between X_1 and X_2 = 0.4, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1					0.3					0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.13	1.09	1.06	1.04	1.02	1.27	1.23	1.12	1.06	0.99	1.44	1.40	1.20	1.09	0.95
	0.75		1.06	1.04	1.03	1.01		1.21	1.11	1.06	1.00		1.38	1.21	1.10	0.98
	1		1.04	1.02	1.01	1.00		1.19	1.11	1.06	1.00		1.36	1.21	1.12	1.00
0.3	0.5	1.27	1.13	1.12	1.11	1.11	1.44	1.29	1.19	1.14	1.08	1.63	1.48	1.29	1.18	1.05
	0.75		1.06	1.06	1.06	1.06		1.22	1.14	1.09	1.04		1.40	1.24	1.14	1.03
	1		0.99	0.99	1.00	1.00		1.14	1.08	1.04	1.00		1.32	1.18	1.10	1.00
0.5	0.5	1.44	1.20	1.20	1.21	1.21	1.63	1.38	1.29	1.24	1.18	1.86	1.61	1.41	1.30	1.17
	0.75		1.06	1.09	1.10	1.12		1.24	1.18	1.14	1.10		1.46	1.30	1.20	1.09
	1		0.92	0.95	0.98	1.00		1.09	1.05	1.03	1.00		1.30	1.17	1.09	1.00

** OR=odds ratio, ICC=intra-class correlation coefficient, α=ln0.1, β_E=0, β_1=β_2=ln2. Width of 95% simulation intervals<0.0053.*

A1.2. Additional tables for four confounders

Table A1.5: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.1, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 Crude OR	0.1 OR adjusted for Z_1	0.1 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 … 0.75	0.1 … 1	0.3 Crude OR	0.3 OR adjusted for Z_1	0.3 … 0.5	0.3 … 0.75	0.3 … 1	0.5 Crude OR	0.5 OR adjusted for Z_1	0.5 … 0.5	0.5 … 0.75	0.5 … 1
0.1	0.5	1.72	1.69	1.67	1.66	1.65	1.94	1.92	1.80	1.73	1.66	2.21	2.20	2.02	1.91	1.79
	0.75		1.68	1.66	1.65	1.64		1.91	1.79	1.73	1.66		2.20	2.01	1.91	1.79
	1		1.67	1.65	1.64	1.63		1.90	1.78	1.72	1.65		2.19	2.01	1.90	1.78
0.3	0.5	1.94	1.81	1.80	1.79	1.78	2.21	2.08	1.95	1.88	1.81	2.55	2.41	2.22	2.10	1.98
	0.75		1.75	1.73	1.73	1.72		2.01	1.88	1.81	1.74		2.33	2.14	2.03	1.90
	1		1.68	1.66	1.66	1.65		1.93	1.81	1.74	1.67		2.25	2.06	1.94	1.82
0.5	0.5	2.21	2.02	2.02	2.01	2.01	2.55	2.37	2.22	2.14	2.06	2.99	2.81	2.60	2.48	2.34
	0.75		1.91	1.91	1.91	1.90		2.25	2.10	2.03	1.94		2.70	2.48	2.35	2.20
	1		1.79	1.79	1.79	1.78		2.13	1.98	1.90	1.82		2.58	2.34	2.20	2.04

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0076.*

Table A1.6: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.2, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.67	1.64	1.62	1.61	1.60	1.87	1.85	1.72	1.64	1.57	2.11	2.11	1.88	1.75	1.60
	0.75		1.63	1.61	1.61	1.60		1.85	1.72	1.64	1.57		2.11	1.89	1.76	1.62
	1		1.62	1.60	1.60	1.59		1.84	1.71	1.65	1.58		2.11	1.90	1.78	1.64
0.3	0.5	1.87	1.72	1.72	1.72	1.71	2.11	1.95	1.82	1.75	1.67	2.40	2.24	2.01	1.87	1.72
	0.75		1.64	1.64	1.64	1.65		1.87	1.75	1.68	1.61		2.16	1.93	1.80	1.66
	1		1.56	1.57	1.57	1.58		1.79	1.67	1.61	1.55		2.07	1.85	1.73	1.59
0.5	0.5	2.11	1.86	1.88	1.89	1.90	2.40	2.15	2.01	1.93	1.85	2.78	2.51	2.25	2.10	1.94
	0.75		1.72	1.75	1.76	1.78		2.00	1.87	1.80	1.73		2.36	2.10	1.96	1.79
	1		1.57	1.60	1.62	1.64		1.84	1.72	1.66	1.59		2.19	1.94	1.79	1.64

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0056.*

Table A1.7: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.3, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2: 0.1 Crude OR	0.1 OR adjusted for Z_1	0.1 OR adjusted for Z_1 and Z_2, ICC of Z_2 = 0.5	0.1 … = 0.75	0.1 … = 1	0.3 Crude OR	0.3 OR adjusted for Z_1	0.3 … and Z_2, ICC of Z_2 = 0.5	0.3 … = 0.75	0.3 … = 1	0.5 Crude OR	0.5 OR adjusted for Z_1	0.5 … and Z_2, ICC of Z_2 = 0.5	0.5 … = 0.75	0.5 … = 1
	0.5		1.60	1.59	1.58	1.57		1.80	1.65	1.57	1.49		2.03	1.78	1.62	1.45
0.1	0.75	1.62	1.59	1.58	1.57	1.57	1.81	1.79	1.66	1.58	1.50	2.02	2.04	1.80	1.66	1.49
	1		1.58	1.57	1.57	1.56		1.79	1.67	1.60	1.52		2.05	1.83	1.70	1.54
	0.5		1.64	1.65	1.66	1.67		1.86	1.72	1.65	1.57		2.11	1.85	1.70	1.54
0.3	0.75	1.80	1.56	1.57	1.58	1.60	2.02	1.77	1.65	1.58	1.51	2.28	2.02	1.78	1.64	1.49
	1		1.47	1.49	1.50	1.52		1.67	1.57	1.51	1.45		1.92	1.71	1.58	1.44
	0.5		1.73	1.78	1.80	1.83		1.98	1.85	1.78	1.71		2.29	2.02	1.86	1.68
0.5	0.75	2.02	1.57	1.62	1.66	1.70	2.28	1.81	1.70	1.64	1.58	2.61	2.11	1.86	1.71	1.55
	1		1.39	1.45	1.49	1.54		1.62	1.54	1.49	1.44		1.91	1.68	1.55	1.40

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0051.*

Table A1.8: Geometric means of the estimated exposure-outcome odds ratios for the case with four confounders for a single standard deviation increase in E when adjusting for Z_1 alone, or Z_1 and Z_2, according to the correlation between confounders and exposure and amount of measurement error in the confounders. Correlation between pairs of confounders = 0.4, ICC of exposure=0.5, sample size=500,000, repetitions=50.

Correlation between E and X_1	ICC of Z_1	Correlation between E and X_2 = 0.1					Correlation between E and X_2 = 0.3					Correlation between E and X_2 = 0.5				
		Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2			Crude OR	OR adjusted for Z_1	OR adjusted for Z_1 and Z_2		
				ICC of Z_2					ICC of Z_2					ICC of Z_2		
				0.5	0.75	1			0.5	0.75	1			0.5	0.75	1
0.1	0.5	1.58	1.57	1.56	1.55	1.55	1.75	1.75	1.60	1.52	1.42	1.95	1.97	1.69	1.53	1.33
	0.75		1.56	1.55	1.55	1.55		1.75	1.62	1.54	1.45		1.98	1.74	1.58	1.39
	1		1.55	1.55	1.55	1.55		1.76	1.64	1.56	1.48		2.01	1.79	1.64	1.47
0.3	0.5	1.75	1.58	1.60	1.62	1.64	1.95	1.77	1.64	1.57	1.49	2.18	2.00	1.74	1.58	1.39
	0.75		1.49	1.52	1.54	1.56		1.68	1.57	1.51	1.44		1.91	1.67	1.53	1.36
	1		1.39	1.42	1.45	1.48		1.58	1.49	1.44	1.38		1.80	1.60	1.48	1.33
0.5	0.5	1.95	1.63	1.69	1.74	1.79	2.19	1.85	1.74	1.67	1.60	2.47	2.12	1.84	1.68	1.49
	0.75		1.45	1.53	1.58	1.64		1.66	1.58	1.53	1.48		1.91	1.68	1.54	1.38
	1		1.24	1.33	1.39	1.47		1.44	1.39	1.36	1.33		1.69	1.49	1.38	1.24

** OR=odds ratio, ICC=intra-class correlation coefficient, correlation between X_3 and E=0.5, correlation between X_4 and E = 0.3, α=ln0.1, β_E=0, $\beta_1=\beta_2=\beta_3=\beta_4$=ln2. Width of 95% simulation intervals<0.0051.*

BLANK IN ORIGINAL

Appendix 2.

Additional graphs from the analyses in Chapter 7

Figure A2.1: Graphical representation of sensitivity analysis results for the effect of a unit increase in FEV_1, a unit increase in log triglycerides, or a doubling of CRP on incident CHD, allowing for measurement error in FEV_1, triglycerides and CRP respectively.

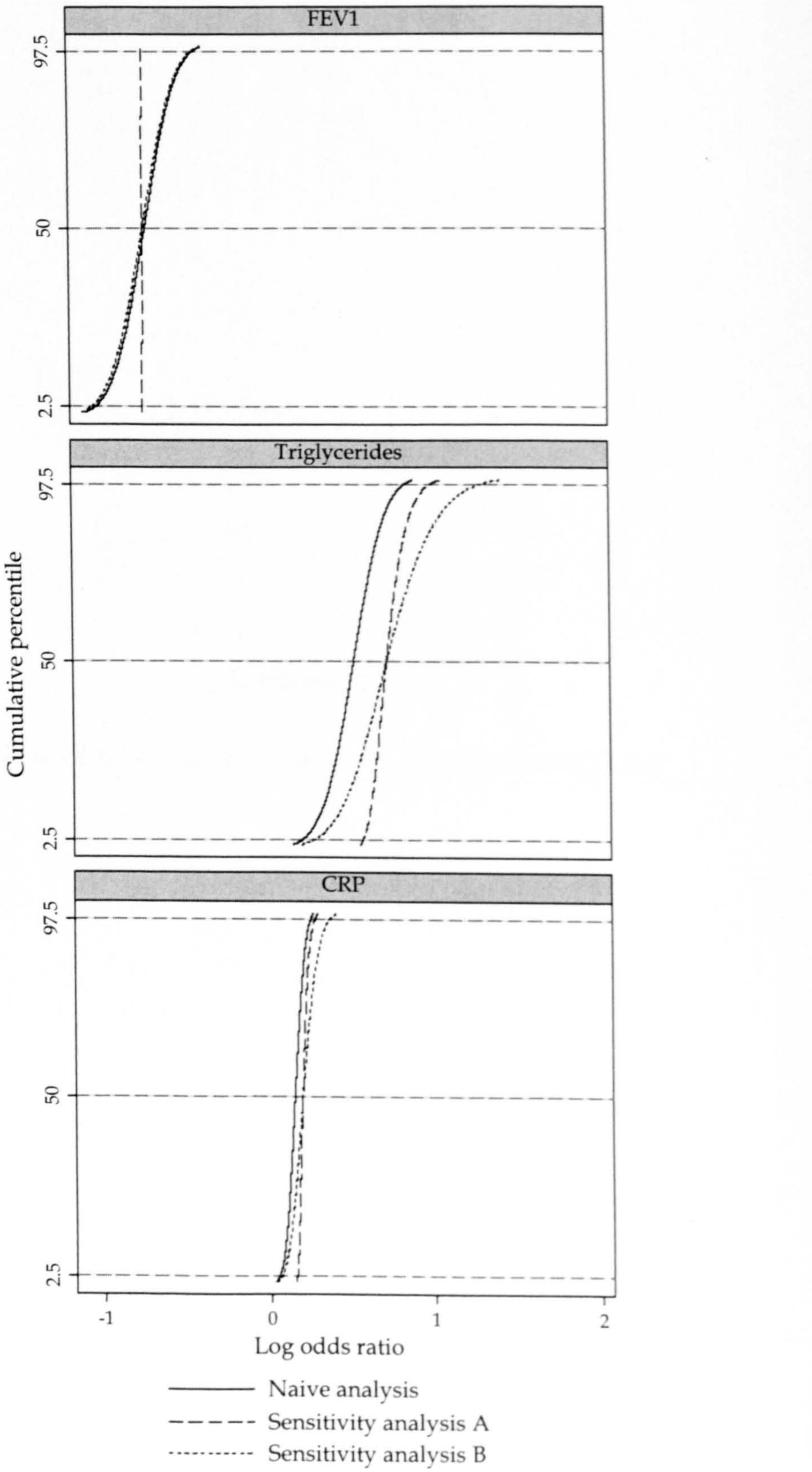


Figure A2.2: Graphical representation of sensitivity analysis results for the effect of a unit increase in FEV_1, a unit increase in log triglycerides, or a doubling of CRP on time to incident CHD, allowing for measurement error in FEV_1, triglycerides and CRP respectively.

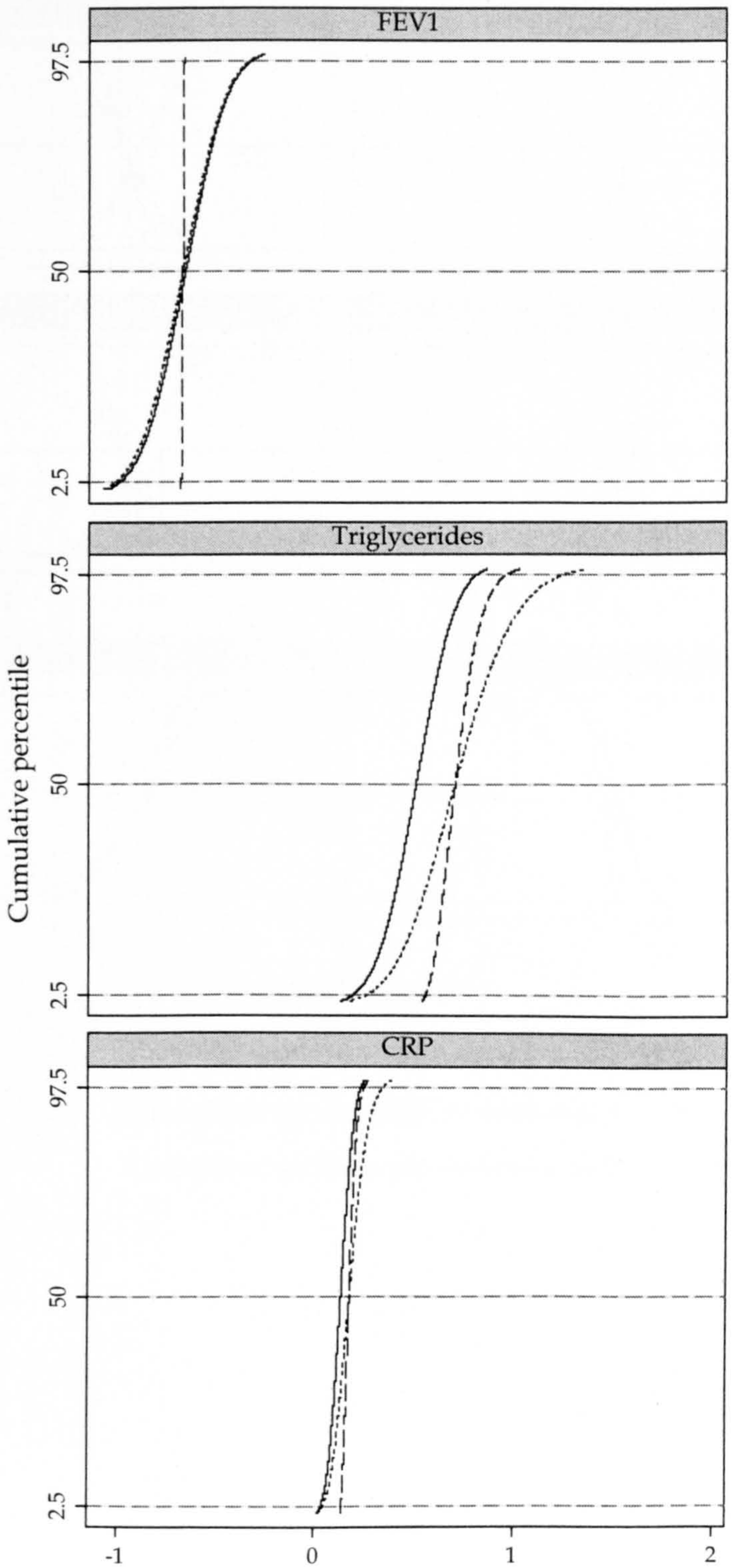


Figure A2.3: Graphical representation of sensitivity analysis results for the effect of doubling CRP on prevalent CHD, allowing for measurement error in FEV_1, triglycerides, CRP or a combination of these.

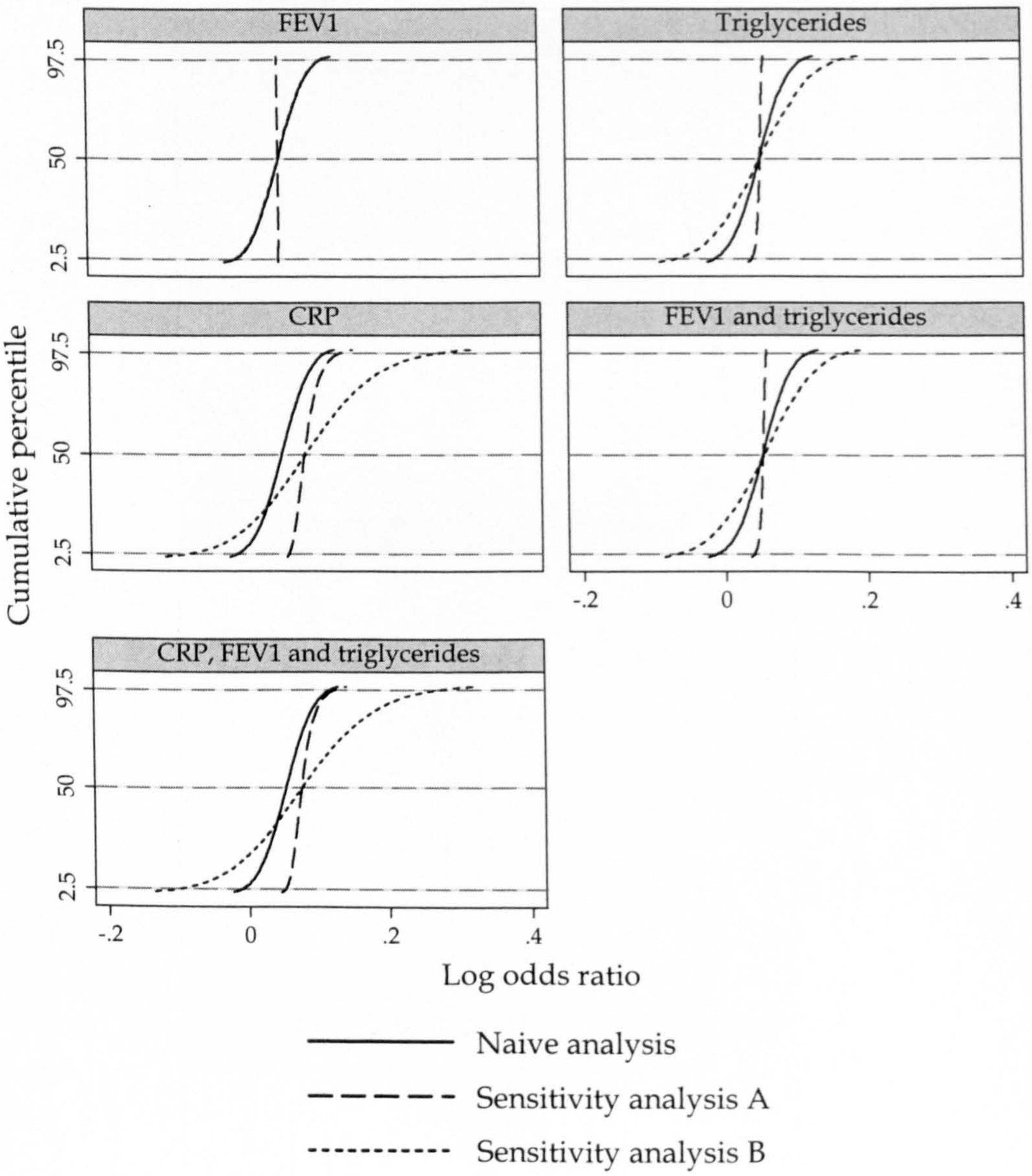


Figure A2.4: Graphical representation of sensitivity analysis results for the effect of doubling CRP on incident CHD, allowing for measurement error in FEV_1, triglycerides, CRP or a combination of these.

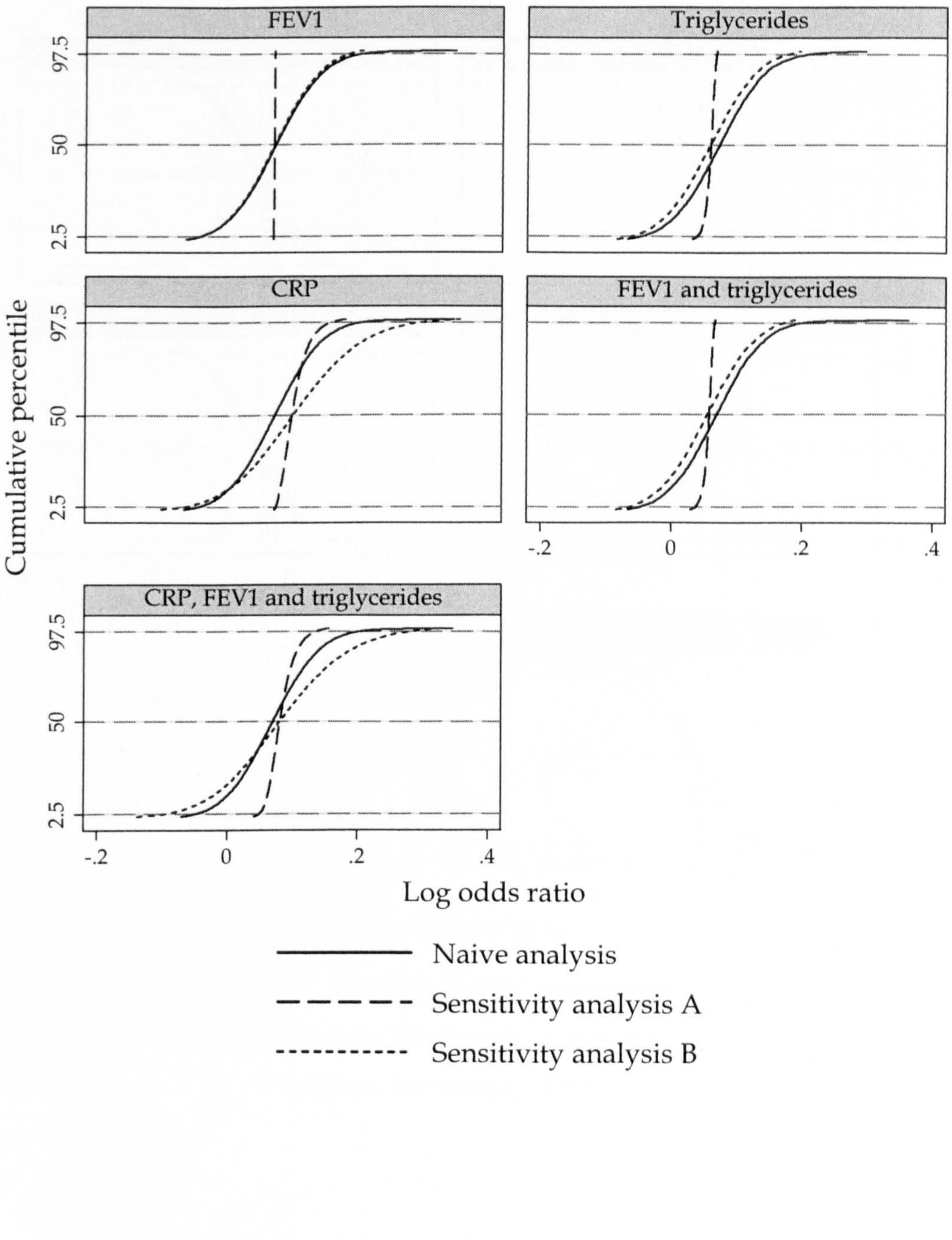


Figure A2.5: Graphical representation of sensitivity analysis results for the effect of doubling CRP on time to incident CHD, allowing for measurement error in FEV_1, triglycerides, CRP, or a combination of these.